

DUE DATE SLIP**GOVT. COLLEGE, LIBRARY**

KOTA (Raj)

Students can retain library books only for two weeks at the most

BORROWER S No	DUE DTATE	SIGNATURE

सांख्यिकी के सिद्धान्त और उपयोग

सांख्यिकी के सिद्धान्त और उपयोग

लेखक

श्री विनोदकरण सेठी

प्रकाशन शाखा, सूचना विभाग

१ -

उत्तर प्रदेश

प्रथम संस्करण

१९६१

मूल्य

९ रुपये

मुद्रक

प० पृथ्वीनाथ भार्गव,

भार्गव श्रृणु श्रेष्ठ, गण्डघाट, वाराणसी

प्रकाशकीय

सांख्यिकी अपेक्षाकृत एक आधुनिक शास्त्र है जिसका महत्त्व ज्ञान-विज्ञान की उन्नति एवं आर्थिक और औद्योगिक समस्याओं की जटिलताओं के साथ बढ़ता जा रहा है। उसके उपयोग का क्षेत्र आज इतना व्यापक हो गया है कि विज्ञान की शायद ही ऐसी कोई शाखा हो जिसमें सांख्यिकी के नियमों और उसके आधार पर प्राप्त तथ्यों का प्रयोग न किया जाता हो। इस समय देश में राद्योत्पादन तथा अन्य वस्तुओं के निर्माण सम्बन्धी जो योजनाएँ बनायी जा रही हैं, उनकी बुनियाद हमारी वर्तमान और भावी आवश्यकताओं तथा वस्तुओं की उपलब्धि सम्बन्धी उन आंकड़ों पर ही रखी जा सकती है जो सांख्यिकी के सिद्धान्तों का सावधानी से प्रयोग करने पर प्राप्त होते हैं। इसी तरह औद्योगिक, आर्थिक तथा चिकित्साविज्ञान सम्बन्धी गवेषणाओं में भी सांख्यिकी द्वारा प्राप्त निष्कर्षों से बड़ी सहायता मिलती है। इसकी इन उपयोगिता और बढ़ते हुए महत्त्व को दृष्टि में रखकर ही यह पुस्तक हिन्दी में प्रकाशित की जा रही है।

हिन्दी-समिति-ग्रन्थमाला की यह ४५वीं पुस्तक है। इसके लेखक श्री विनोद-करण सेठी एम० एस सी० आगरा विश्वविद्यालय के इस्टोर्ट्यूट ऑफ सोशल साइंसेज में सांख्यिकी के सहायक प्राध्यापक हैं। आपने उदाहरण दे-देकर विषय को समझाने की चेष्टा की है जिससे उसकी दुरूहता बहुत घट गयी है।

अपराजिता प्रसाद सिंह

सचिव, हिन्दी समिति

विषय-सूची

भाग एक

परिचय और परिभाषाएँ

	पृष्ठ संख्या
अध्याय १—साहित्यकी क्या है ...	१
११ वैज्ञानिक विधि और साहित्यकी १, १२ साहित्यकी के उपयोग ४।	
अध्याय २—समष्टि और उसका विवरण ...	१३
२१ समष्टि १३, २२ चर १३, २३ आंकड़ों को सक्षिप्त रूप में रखने की विधि १४, २४ आंकड़ों का रेखा चित्रों द्वारा निरूपण १६, २५ चर के परास का विभाजन १९, २६ केन्द्रीय प्रवृत्ति के कुछ माप २५, २७ प्रसार के कुछ माप ३१, २८ घूर्ण ३७, २९ वैषम्य और कठुदता ३८।	
अध्याय ३—प्रायिकता ...	४३
३१ वे स्थितियाँ जिनमें प्रायिकता का प्रयोग किया जाता है ४३, ३२ आपेक्षिक वारम्भारता का सीमान्त मान ४४, ३३ एक अन्य परिभाषा ४६, ३४ प्रतिबन्धी प्रायिकता ४९, ३५ स्वतन्त्र घटनाएँ ५०, ३६ घटनाओं का सगम और प्रतिच्छेद ५०, ३७ परस्पर अपवर्जी घटनाएँ ५१, ३८ घटनाओं का वियोग ५१, ३९ घटनाओं का गमित होना ५२, ३१० आपेक्षिक वारम्भारता के कुछ गुण ५२, ३११ प्रायिकता के गुण ५४, ३१२ वेज का प्रमेय ६०।	
अध्याय ४—प्रायिकता वटन और यादृच्छिक चर .	६५
४१ यादृच्छिक चर ६५, ४२ असतत वटन ६६, ४२१ यादृच्छिक चर के फलन का वटन ६६, ४२२ द्वि-विमितीय यादृच्छिक चर ६८, ४२३ द्वि-विमितीय चर के फलन का वटन ७०, ४२४ एक	

पार्वीय वटन ७१, ४३ सतत वटन ७२, ४३ १ आयताकार वटन ७६, ४३ २ प्रसामान्य वटन ७६, ४४ सचयी प्रायिकता फलन ७७, ४४ १ सचयी प्रायिकता फलन के गुण ७७, ४५ स्वतन्त्र चर ७९, ४६ प्रायिकता वटन के प्रति समाकरण ८१, ४७ यादृच्छिक चर का प्रत्याशित मान अथवा माध्य ८३, ४८ यादृच्छिक चर के शून्य ८४, ४९ स्वतन्त्र चरा के गुणन फल का प्रत्याशित मान ८४, ४१० चरो के योग का प्रत्याशित मान ८५।

भाग दो

परिकल्पना की जाँच और कुछ महत्त्वपूर्ण प्रायिकता वटन	८७
अध्याय ५—मनोवैज्ञानिक पृष्ठ-भूमि	८९
अध्याय ६—द्विपद वटन	१०२
६१ द्विपद वटन १०२, ६२ द्विपद वटन के उपयोग के कुछ उदाहरण १०३, ६३ द्विपद वटन के कुछ गुण १०७, ६४ द्विपद वटन के लिए सारणी १०९, ६५ एक मनोवैज्ञानिक सिद्धान्त की जाँच में द्विपद वटन का उपयोग ११२।			
अध्याय ७—प्लासो वटन	११५
७१ कुछ परिस्थितियाँ जिनमें प्लासो वटन का उपयोग होता है ११५, ७२ द्विपद वटन का सीमान्त रूप ११६, ७३ वास्तविक वटन का प्लासो वटन द्वारा सन्निकटन ११९, ७४ प्लासो वटन के कुछ गुण १२१, ७५ उदाहरण १२५, ७६ प्लासो वटन की सारणी १२६।			
अध्याय ८—प्रसामान्य वटन	१२८
८१ गणितीय वटनो का महत्व १२८, ८२ प्रसामान्य वटन की परिभाषा १३०, ८३ प्रसामान्य वटन के कुछ महत्त्वपूर्ण गुण १३१, ८४ प्रसामान्य वटन द्विपद वटन का एक सीमान्त रूप १३४, ८५ त्रुटियो का वटन १३७, ८६ भाउस के त्रुटि-वटन की व्युत्पत्ति १३९, ८७ परिकल्पनाओ की जाँच में प्रसामान्य वटन का उपयोग १४४।			

अध्याय ९— X^2 वटन १५०
<p>९१ यादृच्छिक चर के फलन का वटन १५०, ९२ X^2 का वटन १५०, ९३ X^2-चर की परिमाणा १५१, ९४ X^2 वटन के कुछ गुण १५२, ९५ समष्टि को पूर्णरूप से विनिर्दिष्ट करनेवाली परिवर्तननाओं के लिए X^2-परीक्षण १५४, ९६ X^2-वटनो की सारणी १५६, ९७ उदाहरण १५७, ९८ आसजन सौष्ठव का X^2-परीक्षण १६०, ९९ समष्टि को अपूर्ण रूप से विनिर्दिष्ट करनेवाली परिवर्तननाओं के लिए X^2-परीक्षण १६०, १०० गुण साहचर्य के लिए दो स्वतन्त्र प्रतिदर्शों का X^2-परीक्षण १६२, १०१ प्रसामान्य-वटन के प्रसरण संबंधी परिवर्तनना-परीक्षण में X^2-वटन का उपयोग १६९।</p>		
अध्याय १०— t -वटन १७२
<p>१०१ उपयोग १७२, १०२ t-वटन का प्रसामान्य-वटन और X^2-वटन से संबंध १७२, १०३ परिवर्तनना परीक्षण १७३, १०४ उदाहरण १७४, १०५ एक तरफा और दो तरफा परीक्षण १७६, १०६ द्वि प्रतिदर्श परीक्षण १७८, १०७ उदाहरण १८०, १०८ t-परीक्षण पर प्रतिबंध १८२,</p>		
अध्याय ११— F -वटन १८४
<p>१११ F-वटन और X^2-वटन का संबंध १८४, ११२ परिवर्तनना परीक्षण १८५, ११३ उदाहरण १८५।</p>		
अध्याय १२—परिवर्तनना की जाँच के साधारण सिद्धान्त	..	१८७
<p>१२१ जाँच की परिचित विधि की आलोचना १८७, १२२ अस्वीकृति क्षेत्र १८७, १२३ एक तरफा परीक्षण १८८, १२४ विभिन्न निकषो से अलग-अलग निष्कर्ष निकालने की संभावना १८८, १२५ नीमन-पीयरसन सिद्धान्त १९०, १२५१ पहली प्रकार की त्रुटि १९१, १२५२ दूसरी प्रकार की त्रुटि १९१, १२५३ सिद्धान्त १९१, १२६ परीक्षण-सामध्य और उसका महत्त्व १९१, १२६१ परिभाषा १९१, १२६२ उदाहरण १९१, १२६३ अभिन्न और अनभिन्नत-</p>		

परीक्षणों की परिभाषा १९२, १२७ प्राचल वा अवकाश १९२,
 १२८ निराकरणोप परिष्कार-पना १९३, १२९ प्रतिदश और प्रतिदश-
 परिमाण १९३, १२१० स्वीकृति और अस्वीकृति क्षम १९४,
 १२११ प्रथम प्रकार की त्रुटि की प्राधिकता और सामर्थ्य १९४,
 १२१२ तुल्य तथा उत्तम परीक्षण १९४, १२१३ प्रमेय १९५,
 १२१४ ग्राह्य परीक्षण १९६ १२१५ अस्वीकृति प्रवेदा के चुनाव के
 अन्य निकष १९७, १२१६ उदाहरण १९७, १२१७ कुछ परि-
 भाषाएँ १९८, १२१८ उदाहरण २००, १२१९ नीमन-पीयरसन के
 सिद्धान्तों की आलोचना २०१, १२२० फिगर की विचारधारा २०२।

भाग तीन

साहचर्य समाश्रयण और सहसंबंध

२०९

अध्याय १३—साहचर्य

...

...

२११

१३१ परिचय २११, १३२ साहचर्य की परिभाषा २१२, १३३
 साहचर्य के माप २१३, १३४ क्रमिक साहचर्य का सूचकांक २१७,
 १३५ क्रमिक साहचर्य के सूचकांक का कलन २१७।

अध्याय १४—सह-संबंध

..

...

२२१

१४१ परिचय २२१, १४२ सह-संबंध सारणी २२१, १४३ धनात्मक
 व ऋणात्मक सह-संबंध २२२, १४४ प्रकीर्ण बिन्दु २२३, १४५
 समाश्रयण बन्ध २२३ १४६ सह-संबंध गुणांक २२४, १४७ समा-
 श्रयण गुणांक और सह-संबंध गुणांक में संबंध २२६, १४८ सह-संबंध
 गुणांक का परिकलन २२७, १४९ बहुत बड़े प्रतिदश के लिए सह-
 संबंध गुणांक का परिकलन २२८, १४९१ परिकलन की जाँच २२८,
 १४१० मूल बिंदु व मानक का परिवर्तन २२९।

अध्याय १५—ब्रह्म-आसन्न

..

.

२३२

१५१ अनुमान में त्रुटि २३२, १५२ अनुमान के लिए प्रतिरूप का
 उपयोग २३४, १५३ अवकल कलन के कुछ सूत्र २३४, १५४ एक-

घात प्रतिरूप का आमजन २३५, १५५ अविक सरल प्रतिरूप २३८,
१५६ प्राक्कलकी के प्रसरण २३९, १५७ परिवर्तना परीक्षण २४१,
१५८ द्विघाती परबलय का आमजन २४२।

अध्याय १६—प्रतिबन्धी वंश, सहसंबंधानुपात और माध्य वर्ग आसंग ... २४५
१६१ असतत चर २४५, १६२ सतत चर २४६, १६३ समाश्रयण
२४८, १६४ सहसंबंधानुपात २४९, १६५ माध्य वर्ग आसंग २५०।

भाग चार

प्राक्कलन

२५३

अध्याय १७—प्राक्कलन के आरम्भिक सिद्धान्त ... २५५
१७१ प्राक्कलक और उसके कुछ इच्छित गुण २५५, १७२ दो अन-
भिन्नत प्राक्कलको का सचयन २५९; १७३ प्राक्कलक प्राप्त करने
की कुछ विधियाँ २६०; १७४ विश्वास्य अंतराल २६५।

भाग पाँच

प्रयोग अभिकल्पना

२६९

अध्याय १८—संपरीक्षण में सांख्यिकी का स्थान ... २७१
१८१ भौतिकी और रसायन के प्रयोगों में सांख्यिकी का साधारण-सा
महत्त्व २७१, १८२ विज्ञान की अन्य शाखाओं में सांख्यिकी का असा-
धारण महत्त्व २७१, १८३ परिकल्पना की जाँच और प्राचलो के
प्राक्कलन में प्रयोग अभिकल्पना का महत्त्व २७२, १८४ उदाहरण
२७३; १८५ यादुच्छिकीकरण २७४, १८६ नियंत्रित यादुच्छिकी-
करण २७६, १८७ ब्लॉक २७७, १८८ प्रयोग आरम्भ करने से पूर्व
योजना की आवश्यकता २७७, १८९ प्रयोग की योजना बनाने समय
तीन बातों का ध्यान रखना होता है २७८, १८१० प्रयोग का उद्देश्य
२७८; १८११ प्रायोगिक उपनार २७९, १८१२ बहु-उपादानीय
प्रयोग २७९; १८१३ नियंत्रण इकाइयों २८०, १८१४ प्रयोग अभि-
कल्पना का एक सरल उदाहरण २८१, १८१५ निराकरणोप परि-
कल्पना को सिद्ध नहीं किया जा सकता २८३, १८१६ भौतिक

स्थितियों पर नियंत्रण की आवश्यकता २८३, १८.१७ प्रयोग को अधिक सुग्राही बनाने के कुछ तरीके २८३ ।

अध्याय १९—प्रसरण-विश्लेषण ... २८६

१९१ एक प्रयोग २८६, १९२ प्रसरणों का संयोज्यता गुण २८६,
 १९३ औसत लम्बाई का प्राक्कलन २८७, १९४ औसत लम्बाई के प्राक्कलक का प्रसरण २८८, १९५ प्रसरण का प्राक्कलन २८९,
 १९५ 10^{-2} का प्राक्कलन २८९ १९५२ 10^{-2} का प्राक्कलन २९०, १९६ प्रसरण विश्लेषण २९१, १९७ प्रसरण विश्लेषण का परिकल्पना की जाँच में उपयोग २९२, १९८ प्रसरण विश्लेषण सारणी २९३, १९९ कुछ कल्पनाएँ जिनके आधार पर निराकरणिय परिकल्पना की जाँच की जा सकती है २९४, १९१० F-परीक्षण २९५ ।

अध्याय २०—यादृच्छिकीकृत ब्लॉक अभिकल्पना ... २९७

२०१ ब्लॉक बनाने का उद्देश्य २९७, २०२ यादृच्छिकीकरण और पुन प्रयोग २९८, २०३ यादृच्छिकीकृत ब्लॉक अभिकल्पना और पूर्णत यादृच्छिकीकृत अभिकल्पना में अन्तर २९८, २०४ वे उपादान जिन पर पैदावार निर्भर करती है ३००, २०५ यादृच्छिकीकृत ब्लॉक अभिकल्पना के लिए एक गणितीय प्रतिरूप ३००, २०६ विभिन्न परिकल्पनाओं के अन्तर्गत 10^{-2} का प्राक्कलन ३०१, २०७ बिना परिकल्पना के 10^{-2} का प्राक्कलन ३०३, २०८ प्रसरण विश्लेषण सारणी ३०३, २०९ परिकल्पनाओं की जाँच ३०५, २०१० उदाहरण ३०५, २०११ ब्लॉक ३०९ ।

अध्याय २१—लैटिन बर्ग अभिकल्पना ... ३१०

२११ प्रयोग को सुग्राही बनाने का प्रयत्न ३१०, २१२ उदाहरण ३१०, २१३ आँकड़े ३१२, २१४ लैटिन बर्ग ३१२, २१५ विश्लेषण ३१३, २१६ साधारण ३१६ ।

अध्याय २२—बहु-उपादानीय प्रयोग ... ३१७

२२१ परिचय ३१७, २२२ बहु-उपादानीय प्रयोग के लाभ ३१८, २२३ मुख्य प्रभाव और परस्पर क्रिया ३१९, २२४ उदाहरण ३२२; २२५ विश्लेषण ३२३ ।

अध्याय २३—समानुलन	३२८
२३ १ असपूर्ण ब्लॉक अभिकल्पना की आवश्यकता ३२८, २३ २ परस्पर क्रिया का समानुलन ३२९, २३ ३ विश्लेषण ३३०, २३ ४ आशिय समानुलन ३३५, २३ ५ सांख्यिकीय विश्लेषण ३३६।			
अध्याय २४—संतुलित असपूर्ण ब्लॉक अभिकल्पना		..	३३८
२४ १ परिभाषा ३३८, २४ २ उदाहरण ३३८, २४ ३ संतुलित असपूर्ण ब्लॉक अभिकल्पना के प्राचलो के कुछ सवय ३४०, २४ ४ यादृच्छिकीकरण ३४१, २४ ५ खेती से मवधित एक संतुलित-असपूर्ण ब्लॉक अभिकल्पना ३४१, २४ ५ १ विश्लेषण के लिए प्रतिरूप, प्रतिरूप के प्राचलो का प्राक्कलन ३४१, २४ ५ २ परिकल्पना परीक्षण ३४३, २४ ५ ३ आँकडे ३४४, २४ ५ ४ विश्लेषण ३४५।			
अध्याय २५—सहकारी घर का उपयोग और सह-प्रसरण विश्लेषण	...		३४७
२५ १ प्रयोग को अधिष दस बनाने का प्रयत्न ३४७, २५ २ समाश्रयण प्रतिरूप ३४७, २५ ३ उपचारो के प्रभाव समान होने की परिकल्पना के अन्तर्गत समाश्रयण प्रतिरूप के प्राचलो का प्राक्कलन ३४८, २५ ४ बिना परिकल्पना के समाश्रयण प्रतिरूप के प्राचलो का प्राक्कलन ३४९, २५ ५ उपचार वर्ग-योग ३५१, २५ ६ परिकल्पनाओ के परीक्षण ३५४, २५ ७ उदाहरण ३५४, २५ ७ १ प्रेक्षण ३५५।			

भाग छ.

प्रतिदर्श सर्वेक्षण

अध्याय २६—प्रतिदर्श सर्वेक्षण के साधारण सिद्धान्त	३६१
२६ १ योजना के लिए सर्वेक्षण की आवश्यकता ३६१, २६ २ सर्वेक्षण में त्रुटियाँ ३६२, २६ ३ अन्य उपादान ३६३, २६ ४ सरल यादृच्छिक प्रतिचयन ३६४, २६ ५ प्राक्कलन ३६५, २६ ६ प्राक्कलक का प्रसरण ३६६, २६ ७ प्राक्कलक के प्रसरण का प्राक्कलन ३६७, २६ ८ अनुपात का प्राक्कलन ३६८, २६ ९ विचरण-गुणाक और प्रतिदर्श परिमाण ३६९।			

अध्याय २७—स्तरित प्रतिचयन	...	३७१
२७ १ परिचय ३७१, २७ २ प्राक्कलन ३७१, २७ ३ प्राक्कलन का प्रसरण ३७२, २७ ४ प्रसरण का प्राक्कलन ३७२, २७ ५ विभिन्न स्तरों में प्रतिदर्श परिमाण का विवरण ३७३, २७ ५ १ समानुपाती वितरण ३७३, २७ ५ २ अनुकूलतम वितरण ३७४, २७ ६ स्तरण-विधि ३७५, २७ ७ सन्निकटन ३७६ ।		
अध्याय २८—द्वि-चरणी प्रतिचयन	...	३७७
२८ १ प्रतिचयन विधि और व्यय ३७७, २८ २ द्वि-चरणी प्रतिचयन विधि ३७७, २८ ३ सकेत ३७८, २८ ४ प्रतिचयन ३७८, २८ ५ प्राक्कलन ३७८, २८ ६ प्राक्कलक प्रसरण ३७९, २८ ७ प्रसरण का प्राक्कलन ३८०, २८ ८ अनुकूलतम वितरण ३८१, २८ ९ उदाहरण ३८३ ।		
अध्याय २९—सामूहिक प्रतिचयन	...	३८५
२९ १ सामूहिक प्रतिचयन ३८५, २९ २ अनुपाती प्राक्कलन ३८५, २९ ३ व्युत्पन्न प्रतिचयन ३८६, २९ ४ प्रारोहक समूह ३८७, २९ ५ सामूहिक प्रतिचयन में प्रसरण ३८८, २९ ६ प्रसरण का प्राक्कलक ३८८, २९ ७ सामूहिक और सरल यादृच्छिक प्रतिचयन की तुलना ३८८ ।		
अध्याय ३०—अनुपाती प्राक्कलन	...	३९०
३० १ अनुपात का प्राक्कलन ३९०, ३० २ अनुपाती प्राक्कलक अभिनति ३९०, ३० ३ अभिनति का प्राक्कलन ३९२, ३० ४ अनुपाती प्राक्कलन की माध्य-वर्ग-त्रुटि ३९२, ३० ५ सषष्टि-योग का अनुपाती प्राक्कलन ३९२, ३० ६ अनुपाती प्राक्कलन और साधारण अतभिनत प्राक्कलन की तुलना ३९३, ३० ७ उदाहरण ३९४, ३० ८ प्रतिदर्श परिमाण ३९४ ।		
अध्याय ३१—विभिन्न-प्राधिकता प्रचयन	...	३९६
३१ १ चयन विधि ३९६, ३१ २ विकल्प विधि ३९८, ३१ ३ प्राक्कलन ३९९, ३१ ४ प्राक्कलक का प्रसरण ३९९, ३१ ५ मापानुपाती प्राधिकता ४००, ३१ ६ प्राक्कलक के प्रसरण का प्राक्कलन ४००, ३१ ७ उदाहरण ४०१, ।		
पारिभाषिक शब्दावली	...	४०५

चित्र-सूची

चित्र सख्या	पृष्ठ सख्या
१—सचयी बारवारता	१७
२—आवृत्ति बहुभुज	१७
३—आयत चित्र	१८
४—उत्तर प्रदेश के पुरपो की आयु-आवृत्ति का आयत चित्र	२०
५—उत्तर प्रदेश में प्रतिशत साक्षरता	२२
६—उत्तर प्रदेश में साक्षरता का आयत चित्र	२२
७—फरीदाबाद के परिवारो का मासिक व्यय के अनुसार वितरण- आयत चित्र	२३
८—फरीदाबाद के परिवारो का मासिक व्यय के अनुसार सचयी आवृत्ति चित्र	२४
९—भारतीय ग्राम परिवारो का अघट्ट क्षेत्रफल के अनुसार वितरण —सचयी आवृत्ति चित्र का एक भाग	२५
१०—असममित तथा सममित वितरण	४०
११—ऊर्ध्व रेखा पर निश्चाना वाघकर चलायी हुई गोलिया का वितरण	४५
१२—चौकी पर कर्पा बिन्दुओ की प्रायिकता	४८
१३—पासा फेंकने पर ऊपर की बिन्दुओ की सख्या का प्रायिकता वटन	६७
१४—एक पासे के छ मुख	६८
१५—चित्र १४ में दिय हुए पासे को फेंकने से प्राप्त द्वि विमितीय घर का वटन	६९
१६—चित्र १४ में दिये हुए पासे को फेंकने से प्राप्त ऊपर के मुख की सख्याओ के योग $(x+y)$ का प्रायिकता वटन	७०
१७—चित्र १५ में दिय हुए प्रायिकता वटन का निर्देशाक्षो पर विक्षप X और Y का एक-माश्वीय वटन	७१
१८—एक सतत वटन का आवृत्तिफलन— $y=f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$	७५
१९—आयताकार वटन में $P[a' < x \leq b]$	७६

चित्र सख्या	पृष्ठ सख्या
२०—आयताकार बटन का सचित प्रायिकता फलन	७८
२१—दो स्वतन्त्र यादृच्छिक चरो के समुक्त और एन-पारश्वीय बटन	८०
२२—एक पासे के छ मुख	८१
२३—चित्र २२ मे दिये पासे को फेंकने से प्राप्त ऊपर की मरुपाओ का समुक्त बटन	८१
२४—	८२
२५— $N(\mu=0)$ का घनत्व-फल	१३३
२६—द्विपद $(\frac{1}{2}, \frac{1}{2})$ का दड चित्र	१३४
२७—द्विपद $(\frac{2}{3}, \frac{1}{3})$ का दड चित्र	१३५
२८—द्विपद $(\frac{4}{5}, \frac{1}{5})$ का दड चित्र	१३६
२९—द्विपद $(\frac{6}{7}, \frac{1}{7})$ का दड चित्र	१३६
३०—द्विपद $(\frac{8}{9}, \frac{1}{9})$ का दड चित्र	१३६
३१—	१९६
३२— $\theta=0$ के एक परीक्षण का सामर्थ्य वक्र	१९८
३३—३५ मे से २० बार भफलता के लिए p का सभावता फलन	२०७
३४—सारणी सख्या १.४.१ के लिए प्रकीर्ण चित्र	२२२
३५—सारणी १.४.२ के लिए प्रकीर्ण चित्र और सरल समाश्रयण रेखा	२३७

कुछ ग्रीक अक्षरों के उच्चारण

α एल्फा	B, β बीटा
Γ गामा	δ डेल्टा
ϵ एप्साइलन	ϕ फाई
χ काई	λ लैम्बडा
μ म्यू	ν न्यू
π पाई	ρ रो
σ टॉ	ψ साई
η ईटा	ξ जिई
θ थीटा	ω ओमेगा
Σ σ सिग्मा	

कुछ गणितीय संकेत

(1) e एक संख्या है जिसका मान निम्नलिखित अनंत श्रेणी से प्राप्त होता है।

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \dots + \frac{1}{r!} + \dots$$

$$= 1 + \sum_{r=1}^{\infty} \frac{1}{r!}$$

(2) π (पाई) एक वृत्त की परिधि और व्यास का अनुपात। इसका मान लगभग 3.14159 होता है।

$$(3) \Gamma(n) = \int_0^{\infty} x^{n-1} e^{-x} dx$$

गामा फंक्शन का निम्नलिखित महत्वपूर्ण गुण होता है

$$\Gamma(n+1) = n \Gamma(n)$$

(4) $a \approx b$ a लगभग b के बराबर है।

(5) 'क' > 'ख' 'ख' से 'क' बड़ा है।

(6) 'क' < 'ख' 'ख' से 'क' छोटा है।

(7) $n!$ n वस्तुओं के कुछ क्रमबद्धों की संख्या।

(8) $\binom{n}{r}$ n वस्तुओं में r वस्तुओं के विभिन्न समूहों की

$$\text{संख्या} = \frac{N!}{r! (N-r)!}$$

(9) $A \cup B$ 'A समूह B' A या B में से कम से कम एक घटना का घटित होना

$$(10) \bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup A_3 \dots \cup A_n \quad i=1 \text{ से लेकर}$$

$i=n$ तक A_i घटनाओं का समूह अर्थात् इन n घटनाओं में से कम से कम एक का घटित होना।

- (11) $A-B$ 'A वियोग B' A घटित हो, परन्तु B नहीं ।
- (12) $A \cap B$ 'A प्रतिच्छेद B' A और B दोनों का एक साथ घटना ।
- (13) $C \subset A$ 'घटना C घटना A में सम्मिलित है' अर्थात् यदि C घटित होगी तो A भी घटित होगी ।
- (14) $C \not\subset A$ 'घटना C घटना A में सम्मिलित नहीं है' मानी यदि C घटित हो तो यह आवश्यक नहीं है कि A भी घटित हो ।
- (15) $v(A)$ न्यू ए 'घटना A की बारंबारता' ।
- (16) $P(A)$ 'घटना A की प्रायिकता' ।
- (17) $P(X=a)$ X के a के बराबर होने की प्रायिकता ।
- (18) $P(a < X \leq b)$ X का मान a से अधिक और b के बराबर अथवा b से कम होने की प्रायिकता ।
- (19) $g^{-1}(a, b)$ X के उन मानों का कुलक जिनके लिए $a < g(X) \leq b$
- (20) $\theta \in \omega$ 'थीटा स्थित है ओमेगा में' अर्थात् कुलक ω के मानों में से θ एक है ।
- (21) $P(A/B)$ 'प्रायिकता A दत्त B' यह दिया होने पर कि B घटित हो चुकी है A की प्रतिवधी प्रायिकता ।
- (22) $f(x)$ 'चर X का x पर प्रायिकता घनत्व'

$$= \lim_{\substack{\delta \rightarrow 0 \\ \delta' \rightarrow 0}} \frac{P[x - \delta < X \leq x + \delta']}{\delta + \delta'}$$
- (23) $F(x)$ 'चर X का x पर संचयी प्रायिकता फलन'

$$= P[X \leq x]$$
- (24) (a, b) उन सख्याओं का कुलक जो a से बड़ी और b से छोटी है ।
- (25) (a, b) उन सख्याओं का कुलक जो a के बराबर या a से बड़ी है और b से छोटी है ।
- (26) (a, b) उन सख्याओं का कुलक जो a से बड़ी है और b के बराबर अथवा b से छोटी है ।
- (27) (a, b) उन सख्याओं का कुलक जो न तो a से छोटी है और न ही b से बड़ी ।

भाग ३

परिचय और परिभाषाएँ



अध्याय १

सांख्यिकी क्या है ?

§ १.१ वैज्ञानिक विधि और सांख्यिकी

“अमुक ब्राड का घी बहुत शुद्ध व उत्तम होता है।”

“अमुक देश के लोग बहुत असम्य और निर्दयी होते हैं।”

“विश्व की ९० प्रतिशत जनसंख्या युद्ध के विरुद्ध है।”

“स्ट्रैप्टोमाइसीन से क्षयरोग में कुछ भी लाभ नहीं होता।”

इस प्रकार के अनेको वक्तव्य आपने अपने जीवन में सुने होंगे। यदि आप इनका विश्लेषण करें तो आपको कई आश्चर्यजनक बातों का पता लगेगा। जिन सज्जनों ने उक्त ब्राड के घी की बहुत प्रशंसा की थी उन्होंने संभवतः उस ब्राड के केवल एक ही टिन का उपयोग किया है, जो बहुत उत्तम था।

उक्त विशिष्ट देश के लोगों से जिनको सिखाया है वे उस देश के दो चार व्यक्तियों को छोड़कर अधिक लोगों के सम्पर्क में नहीं आये हैं। जिनको विश्व की जनता का मत जानने का दावा है वे केवल पत्रकार हैं। दो माह के शीघ्रता में किये गये परिभ्रमण के पश्चात् उन्होंने विश्व की जनता की मानसिक स्थिति की विवेचना करते हुए एक पुस्तक लिखी है। उनसे प्रश्न करने पर आपको ज्ञात होगा कि इस भ्रमण में सौ-डेढ़ सौ से अधिक व्यक्तियों से वार्तालाप करने और उनका मत जानने का उन्हें अवसर नहीं मिला। स्ट्रैप्टोमाइसीन पर जिस महिला को विश्वास नहीं है उसका बेटा इसका इजेक्शन लगाने पर भी क्षयरोग से छुटकारा नहीं पा सका था। ये वक्तव्य इन व्यक्तियों ने अपने अनुभव के आधार पर ही दिये थे। परन्तु इन अनुभवों के विश्लेषण में आप इस निष्कर्ष पर पहुँचेंगे कि इस प्रकार कुछ थोड़े से विशिष्ट अनुभवों के आधार पर इतने व्यापक वक्तव्य देना उचित नहीं है। पर यदि आप स्वयं अपने द्वारा दिये गये किसी व्यापक वक्तव्य का विश्लेषण करें तो आपको आश्चर्य होगा कि उसका आधार भी कुछ गिने-चुने अनुभव ही हैं। परन्तु क्यों कि वे स्वयं आपके अनुभव हैं, इसलिए उन वक्तव्यों में आपको पूर्ण विश्वास

है। क्या यह सम्भव नहीं है कि ऊपर जिन वनतव्यों की विवेचना की गयी है वे सब सही हैं—या उनमें से कुछ सही हों? मान लीजिए कि जिस महिला ने स्ट्रेप्टो-माइसीन की आलोचना की थी उन्होंने उन हजारों धयरोगियों का अध्ययन किया होता जिनको स्ट्रेप्टोमाइसीन दी गयी और उनमें से कोई भी रोग से छुटकारा नहीं पा सका। तो क्या फिर भी आप उनके कथन को अनुचित मानते? लेकिन यह अनुभव भी तो विशिष्ट ही है। उन्होंने उन सब रोगियों का तो अध्ययन नहीं किया जिनको यह औषधि दी गयी है। फिर भी उनके कथन में आपका बिश्वास अवश्य ही अधिक दृढ़ होता।

यह शायद मनुष्य का स्वभाव है कि अपने अनुभवों के आधार पर वह उन बहुत-सी वस्तुओं और घटनाओं के बारे में भी एक धारणा बना लेता है जिनका उसे कुछ भी अनुभव नहीं होता। वास्तव में विज्ञान का विकास इसी प्रकार होता है। जब कोई वैज्ञानिक किसी सिद्धान्त अथवा नियम का प्रतिपादन करता है तो उसका आधार भी उसके या अन्य वैज्ञानिकों के अनुभव ही होते हैं। “लोहे के टुकड़े को पानी में रखने से उसमें जग लग जाना है और सोडियम के टुकड़े को पानी में डालने से उसमें आग लग जाती है।” “प्रत्येक द्रव्य-कण हर दूसरे द्रव्य-कण को आकर्षित करता है।” ‘मलेरिया बुखार एनाफिलीस नामक मच्छर के काटने से ही होता है।’ ये सब इस प्रकार के कथन हैं जिन्हें वैज्ञानिक सत्य की सजा दी जाती है। क्या इनके प्रतिपादन का अर्थ यह है कि वैज्ञानिकों ने प्रत्येक लोहे या सोडियम के टुकड़े को पानी में डालकर देखा है या उन्होंने मलेरिया के प्रत्येक रोगी को मच्छर द्वारा काटे जाते हुए देखा है? इस प्रकार किसी भी वैज्ञानिक नियम की विवेचना यदि आप करें तो आपको पता चलेगा कि उनका आधार कुछ सीमित अनुभव ही है।

इस प्रकार विशिष्ट से व्यापक नियमों के प्रतिपादन में दोनों ही सम्भावनाएँ हैं। वे सत्य भी हो सकते हैं और असत्य भी। वैज्ञानिक इस वास्तविकता को समझता है। वह यह दावा नहीं करता कि ये नियम निरपेक्ष सत्य ही हैं। वह यह जानता है कि ये केवल परिकल्पना (hypothesis) मात्र हैं जो वैज्ञानिक जगत् के अभी तक के अनुभवों को समझने में सहायक होते हैं। यदि इन परिकल्पनाओं के विरुद्ध कुछ भी प्रमाण मिलते हैं तो वह इन नियमों में संशोधन करने के लिए अथवा उन्हें त्याग कर दूसरे नियम प्रतिपादित करने के लिए प्रस्तुत रहता है।

व्यापक ज्ञान प्राप्त करने की एक विधि है जिसे वैज्ञानिक विधि कहा जाता है। इसमें निम्न चरण होते हैं—

(१) प्रथम, वस्तुओं, कार्यों और घटनाओं का प्रेक्षण तथा अध्ययन किया जाता है।

(२) द्वितीय, इन प्रेक्षणों में पारस्परिक सम्बन्ध स्थापित करने और उन्हें समझने के लिए कुछ मिद्धान्तों का प्रतिपादन किया जाता है।

(३) तृतीय, इन नियमों में से कुछ निगमन निकाले जाते हैं जो प्रेक्षणगम्य वस्तुओं तथा घटनाओं से सम्बन्धित होते हैं।

(४) चतुर्थ, इन घटनाओं या वस्तुओं के निरीक्षण के लिए कुछ प्रयोगों का आयोजन किया जाता है।

(५) पंचम, यदि इन प्रयोगों के निष्कर्ष प्रतिपादित नियमों के विरुद्ध होते हैं तो इन नियमों को त्याग कर अथवा उनमें सुधार कर नवीन नियम प्रतिपादित किये जाते हैं।

इस प्रकार निरीक्षण और प्रयोग विज्ञान के अभिन्नतम अंग हैं।

किसी साधारण मनुष्य और वैज्ञानिक में यही अन्तर है कि पहला अपने कथनों की पुष्टि के लिए और अधिक निरीक्षण की आवश्यकता नहीं समझता, जब कि दूसरा परीक्षण को अत्यन्त आवश्यक ही नहीं समझता बल्कि परीक्षण और निरीक्षण के बाद भी कथन के असत्य होने की संभावना से परचित है। दार्शनिक तत्त्व-विद्या (meta-physics) का तर्क विज्ञान में प्रयोग होनेवाले तर्क से एकदम विपरीत होता है। उसमें यदि अनुभव किसी नियम का खण्डन करते पाये जाते हैं तो इसे अनुभवों का दोष समझा जाता है, न कि नियमों का।

इस प्रकार वैज्ञानिक विधि से जो ज्ञान प्राप्त किया जाता है वही विज्ञान है। इसमें दो प्रकार के नियम होते हैं। एक तो वे जो यथार्थ हैं जिनके उदाहरण पहले दिये जा चुके हैं। "सोडियम के टुकड़े को पानी में डालने से उसमें आग लग जाती है" यह नियम सोडियम के प्रत्येक टुकड़े पर हर समय लागू होता है। इसी प्रकार जब यह कहा जाता है कि "एनाफिलीस मच्छर के काटने से ही मलेरिया होता है" तो इस कथन का तात्पर्य यह होता है कि किसी भी मनुष्य को बिना इस मच्छर के काटे हुए मलेरिया नहीं हो सकता। इस प्रकार के सब नियम, जिनमें कोई अपवाद नहीं होता, यथार्थ नियम (exact laws) कहलाते हैं। भौतिकी और रसायन-विज्ञान में बहुधा ऐसे ही नियम पाये जाते हैं। कभी-कभी प्रायोगिक फलों और इन नियमों में कुछ अन्तर पाया जाता है, परन्तु यह अन्तर अधिकतर सूक्ष्म होता है—इतना सूक्ष्म कि इसको प्रयोग सम्बन्धी त्रुटि (experimental error) माना जा सकता है।

इसके विपरीत कई परिस्थितियों में एक ही प्रकार की स्थिति और एक से कारणों के रहते हुए भी अलग अलग अनेकों फल सम्भव हो सकते हैं। हो सकता है कि ऐसे कुछ अज्ञात कारण हों जो इन फलों को निश्चित करते हैं। लेकिन इन कारणों के ज्ञान के अभाव में किसी यथार्थ नियम को प्रतिपादित करना असम्भव है। जैसे यह कहना असम्भव है कि किसी स्त्री की आगामी सन्तान लड़की होगी या लड़का, अथवा स्ट्रैप्टोमाइसीन से कोई विद्येय मरीज नीरोग हो जायगा या नहीं; या किसी निर्दिष्ट ताप, नमी व हवा के स्तर और वेग के होने पर वर्षा होगी या नहीं। ऐसी अवस्था में किसी निर्दिष्ट वस्तु अथवा घटना के बारे में भविष्य वाणी करने में दोनो ही सम्भावनाएँ हैं। ये भविष्य कथन सत्य भी हो सकते हैं और असत्य भी। लेकिन ऐसी परिस्थितियों में भी वैज्ञानिक बिल्कुल विवश नहीं हो जाता। वह यथार्थ से भिन्न एक दूसरे प्रकार के नियम का प्रतिपादन कर सकता है। ये नियम अकेली वस्तुओं अथवा घटनाओं के बारे में नहीं होते बल्कि अनेक एक-सी वस्तुओं अथवा घटनाओं के समुदायों के बारे में होते हैं। ये नियम यह बताते हैं कि इस समुदाय में प्रयोग के फलस्वरूप जो भिन्न भिन्न फल प्राप्त होंगे उनकी बारम्बारता (frequency) कितनी होगी। उदाहरण के लिए "१०० बच्चों में से ५१ लड़कियाँ होती हैं और ४९ लड़के" अथवा "८० प्रतिशत क्षयरोगियों को स्ट्रैप्टोमाइसीन से लाभ होता है।"

ऐसे नियमों को सांख्यिकीय नियम (statistical laws) कहा जाता है। इस प्रकार सांख्यिकी में निम्नलिखित बातें सम्मिलित हैं।

- १—घटनाओं या वस्तुओं के गुणों का सामुदायिक रूप में प्रेक्षण करना।
- २—इन प्रेक्षणों का विश्लेषण करके संक्षिप्त रूप में उनका वर्णन करना।
- ३—इस वर्णन के आधार पर बारम्बारता अथवा प्रायिकता (probability) के रूप में नियमों का प्रतिपादन करना।

४—कुछ दूसरी प्रेक्षणगम्य (observable) घटनाओं की प्रायिकता सम्बन्धी निष्कर्ष निकालना।

५—इन निष्कर्षों की जाँच करने के लिए कुछ प्रयोगों का आयोजन करना।

६—इन प्रयोगों के फलों का विश्लेषण करना।

§ १.२ सांख्यिकी के उपयोग

वे परिस्थितियाँ जिनमें सांख्यिकीय रीति का उपयोग होता है इतनी व्यापक हैं कि विज्ञान की ऐसी शाखा कदाचित् ही कोई हो जिसमें इस रीति का उपयोग कभी

न किया जाता हो। भौतिक तथा रासायनिक विज्ञानों में भी, जिन्हें बहुत समय तक पूर्णतः यथार्थ समझा जाता था, कई नियम प्रायिकताओं के रूप में हैं। विशेषतः इलेक्ट्रान, प्रोटीन और न्यूट्रान आदि सूक्ष्म कणिकाओं के अध्ययन में तो सांख्यिकीय नियमों का ही प्रयोग किया जाता है। जो नियम बड़े पिण्डों के सम्बन्ध में होते हैं वे यथार्थता के इतने निकट होते हैं कि नियम और फलों के अन्तर को प्रायोगिक भूल समाप्त कर उनकी उपेक्षा की जा सकती है। अतः कई वैज्ञानिक यह बात मानने लगे हैं कि वैज्ञानिक नियम कभी भी पूर्ण रूप से यथार्थ नहीं होने बल्कि यथार्थ के सन्निकटन-मात्र होते हैं। ये मानते हैं कि सभी नियमों की प्रकृति अन्तिम विद्वेषण में सांख्यिकीय ही होती है।

आरम्भ में विज्ञानों में सांख्यिकी का उपयोग अधिकतर प्रयोगों के समुदाय को इस प्रकार व्यक्त करने में होता था कि उससे प्रवृत्तियाँ (tendencies) प्रत्यक्ष हो जायें। फिर कुछ विज्ञानों में व्यक्तियाँ और इकाइयों को छोड़कर इनके समूह के आचरण के अध्ययन पर जोर दिया जाने लगा। इसके लिए सांख्यिकीय रीतियाँ बहुत उपयुक्त तथा आवश्यक थीं।

कृषि व प्राणि-विज्ञान के अध्ययन में वैज्ञानिकों को आरम्भ में बहुत अधिक कठिनाई का सामना करना पड़ा था। किन्हीं दो पायों पर एक ही प्रकार की खाद व पानी का एक-मा असर नहीं पड़ता। यही बात पशुओं में भी पायी गयी। ऐसी दशा में एक ही उपाय था। वह यह कि व्यक्ति-वित्तों को छोड़कर उनके समुदायों के नियम में नियमों की खोज की जाय। इस दृष्टिकोण से विद्वेषण की अधिक उन्नत विधियाँ की आवश्यकता पूरी करने में सांख्यिकीय तरीका का प्रयोग हुआ। नयी नयी परिस्थितियों का सामना करने के लिए नये नये सिद्धान्त बनाये गये। इस प्रकार सांख्यिकी के विकास में कृषि एवं प्राणि-विज्ञान का बहुत बड़ा भाग है।

इन विज्ञानों में केवल यही आवश्यकता नहीं थी कि प्रयोगों के फलों की ठीक से विवेचना की जाय। इस व्याख्या को सरल और प्रयोगों को अधिक सफल बनाने के लिए प्रयोगों के आयोजन में भी उन्नति की आवश्यकता थी। किसान यह चाहता है कि अनाज के उत्पादन का स्तर ऊँचा बना रहे। उसकी सहायता के लिए कृषि-विज्ञान वेत्ताओं को प्रयोग करने होते हैं जिनसे यह मालूम हो जाय कि अनाज की भिन्न-भिन्न किस्मों के प्रयोगों से उपज में क्या अन्तर पड़ जाता है, विभिन्न खादों के क्या प्रभाव हैं और खेती करने की सबसे उत्तम विधि क्या है। यह आशा की जाती है कि इन प्रयोगों के आधार पर वह किसानों को लाभदायक सुझाव दे सकेगा।

विभिन्न खादों की तुलना के लिए पहले-पहल जो प्रयोग किये गये थे उनमें यह काफी समझा गया था कि खादों का भिन्न-भिन्न भू-क्षेत्रों में प्रयोग किया जाय और उनके उत्पादन की तुलना करके उनके आपेक्षिक मूल्य का तर्कसम्मत अनुमान लगा लिया जाय। परन्तु धीरे-धीरे अनुसंधान-कर्त्ताओं को पता लग गया कि इस तरीके से समुचित मूल्यांकन होता असंभव है। एक ही विस्म के पौधों की उपज में, जिन्हें भिन्न-भिन्न भू-क्षेत्रों में बोकर एक ही प्रकार की मिट्टी, खाद व पानी का उपयोग किया गया हो, बहुत अन्तर हो सकता है। इसलिए जब खादों की तुलना की जाय तो इस बात का पता चलाना आवश्यक हो जाता है कि जो अंतर उत्पादन में पाया जाता है उसका सबब खादों से ही है अथवा उन अनेक कारणों से जिनसे या तो बैज्ञानिक अनभिज्ञ है या जिन पर उनका कुछ बड़ा नहीं है। इसके लिए सांख्यिकीय तर्कों का प्रयोग किया गया है और बैज्ञानिक अन्वेषण में उसका महत्त्व प्रमाणित हो चुका है।

कृषि-विज्ञान से ही सर्वाधिक वनस्पति-प्रजनन (plant breeding) विज्ञान है। वनस्पति-संवर्धक किन्मी भी गवेषणा का अंतिम ध्येय होता है वनस्पति की अधिक उन्नत किस्मों का विकास। किसी भी किस्म की उन्नति कई विभिन्न दृष्टि-काणों से हो सकती है। उदाहरणार्थ वनस्पतियों को जो खाद दी जाती है वे उसका उपयोग करने के योग्य बनें, बीमारी के कीटाणुओं से वे अधिक सुरक्षित हो या तापमान के उतार-चढ़ाव को सहन करने की उनकी शक्ति में वृद्धि हो। वनस्पति पर उत्पत्ति-संबन्धी और वातावरण-संबन्धी उपादानों (factors) का प्रभाव पड़ता है। जिस प्रकार किसान अनुकूल वातावरण द्वारा अधिक उत्पादन प्राप्त करने की चेष्टा करता है, उसी प्रकार वनस्पति-प्रजनन का अध्ययन करनेवाला उत्पत्ति के सिद्धान्तों के उपयोग द्वारा वनस्पतियों के वशानुगत गुणों में उन्नति करने का प्रयत्न करता है। परन्तु इस गवेषणा में उसे नये नये प्रश्नों को हल करना पड़ता है जिसके लिए वे सिद्धान्त ग्रहण नहीं होते जिनका उसे पहले से ज्ञान है। नये सिद्धान्तों की खोज के लिए उसे उत्पत्ति सम्बन्धी प्रयोग करने पड़ते हैं। इस गवेषणा में जितना धन उपलब्ध है और जितना समय है उसको देखते हुए किस प्रकार पौधों का चुनाव करना चाहिए, प्रयोग के लिए उनकी संख्या किस प्रकार निर्धारित करनी चाहिए, भिन्न-भिन्न श्रेणियों को भिन्न-भिन्न भू-क्षेत्रों में किस नियम के अनुसार लगाना चाहिए आदि समस्याओं का हल सांख्यिकी के सिद्धान्तों के उपयोग से ही होता है।

पिछले दस पन्द्रह वर्षों में विटामिनो के सबब में बहुत अनुसंधान हुआ है। भिन्न-भिन्न विटामिनो के महत्त्व को समझने के लिए अनेक प्रयोग किये गये हैं। यह

प्रयोग बहुधा पशुओं पर किये जाते हैं, क्योंकि उम्र, वजन, लिंग, बल और पहले से बनी हुई भोजन की आदतें आदि कई बातें हैं जो भोजन के प्रभाव को किसी सीमा तक निर्धारित करती हैं, इसलिए इन प्रयोगों के लिए पशुओं के ऐसे समूहों को चुना जाता है जो ऊपर लिखी बातों में एक-से अथवा लगभग एक-से हों। एक समूह को एक निर्दिष्ट मात्रा में सामान्य खुराक दी जाती है। शेष समूहों को विटामिनों की भिन्न-भिन्न मात्राओं से युक्त खुराक दी जाती है। इनमें से एक को उपर्युक्त सामान्य खुराक से कहीं अधिक विटामिन मिलता है और दूसरे को बहुत कम, लगभग नहीं के बराबर। बाकी समूहों को इन सीमाओं के बीच में भिन्न भिन्न मात्राओं का विटामिन मिलता है। किस पशु को किस समूह में रखा जाये यह अनिश्चितता में निश्चित किया जाता है। पशुओं को इन निश्चित खुराकों पर निश्चित समय के लिए रखा जाता है। अन्वेषक प्रतिदिन वजन के उतार-चढ़ाव व वीमारियों के चिह्नों के प्रकट होने का विवरण लिखता रहता है। यदि यह प्रयोग सांख्यिकीय सिद्धान्तों के अनुसार सावधानी से किया गया हो तो इससे कई मूल्यवान् निष्कर्ष निकाले जा सकते हैं।

सामाजिक विज्ञानों में भी सांख्यिकीय विधियाँ का बहुत उपयोग होता है। जनता का मत जानने में राजनीतिक दलों की रुचि होना स्वाभाविक ही है और इस कारण वे सांख्यिकी से अधिक परिचित होते जा रहे हैं। अर्थशास्त्र की गवेषणाओं में तो सांख्यिकीय विधियाँ अपरिहार्य हो जाती हैं। अर्थशास्त्र के नियमों का सवध सामुदायिक प्रवृत्तियों में होता है और ऐसे निष्कर्षों का निर्धारण बहुधा सांख्यिकीय प्रणाली के विवेकपूर्ण उपयोग पर निर्भर करता है।

पोषण-सम्बन्धी गवेषणा (nutritional research) में एक लक्ष्य यह हो सकता है कि भोजन में उन तत्वों का अधिक से अधिक उपयोग हो जिनकी, भोजन संधी अध्ययन के अनुसार, औसत आहार में कमी पायी गयी है। इस क्षेत्र में सांख्यिकी का प्रारम्भिक कार्य केवल प्रति मनुष्य औसत आहार का पता लगाना और उसकी किसी लक्ष्य से तुलना करना है। प्रति व्यक्ति आहार का वितरण किस प्रकार है यह जानना भी उतना ही महत्त्वपूर्ण है जितना औसत का ज्ञान। परन्तु एक बड़े देश में प्रत्येक मनुष्य से उसके आहार का विवरण प्राप्त करना असम्भव सा है। यदि यह सम्भव भी हो तो इन आँकड़ों को जोड़कर उनसे औसत का परिकलन करने में त्रुटि होने की इतनी अधिक सम्भावना है कि इतना अधिक व्यय करके इन आँकड़ों को प्राप्त करना उचित प्रतीत नहीं होता। जिस प्रकार एक गुड़ी चावल का नमूना देखने से

एक बोरे चावल की किस्म का अनुमान लगाया जाता है उसी प्रकार कुछ घोंडे से मनुष्यों को चुनकर और उनके आहार सब्जी आँकड़ों को एकत्र करके क्या देश के औसत का पता नहीं लगाया जा सकता ? सार्विकीय सिद्धान्तों के प्रयोग से यह निर्णय किया जा सकता है कि इस कार्य के लिए कितने मनुष्यों का चुनाव मथेष्ट होगा या उनका चुनाव किस प्रकार किया जाये कि औसत का अनुमान अधिक विश्वसनीय हो।

देश के बारे में साधारण ज्ञान सरकार के लिए बहुत ही आवश्यक होता है। देश में कितना अनाज उत्पन्न हुआ है और कितने अनाज की आवश्यकता है, इसका यदि सरकार को अनुमान न हो तो अनाज के उत्पात नियंत्रण के बारे में किसी निर्णय के लिए उसके पास कोई विश्वसनीय आधार नहीं होगा। यदि उसे यह पता न हो कि देश में उपज आवश्यकता में एक करोड़ टन कम हुई है तो हो सकता है कि उसे अकाल का सामना करना पड़े। यदि अनाज आवश्यकता में अधिक उत्पन्न हो गया और सरकार इस ज्ञान के अभाव में अनाज के निर्यात पर रोक लगा देती है तो अनाज के दाम गिरकर देश में मदी की स्थिति पैदा हो सकती है। विशेष रूप से आजकल सरकार आगामी पाँच या दस वर्षों की योजनाएँ बनाने में लगी हुई है इसलिए उसके लिए इस प्रकार के ज्ञान की आवश्यकता बहुत बढ़ गयी है। यदि सरकार ने यह निर्णय कर लिया है कि पाँच साल में प्रति व्यक्ति को आय में १० प्रतिशत वृद्धि हो जायेगी तो उसे इस बात का भी अनुमान होना चाहिए कि इस बड़ी हुई आय का मनुष्य क्या करेगा। किस वस्तु को माँग कितनी बढ़ेगी और किस वस्तु को गिरेगी। या यदि उसने इरादा किया है कि राष्ट्रीय आय में १५ प्रतिशत की वृद्धि होगी तो उसे यह भी मालूम होना चाहिए कि जनमस्या किस तेजी के साथ बढ़ रही है। हो सकता है कि योजना-काल के अन्त में राष्ट्रीय आय में वृद्धि होते हुए भी प्रति-व्यक्ति औसत आय में कमी हो जाये। इस प्रकार का साधारण ज्ञान प्राप्त करने के लिए सर्वेक्षण (survey) की आवश्यकता पड़ती है। परन्तु यदि इसके लिए प्रत्येक मनुष्य से पूछनाछ की जाये तो ही संभ्रता है सरकार की सारी आय सर्वेक्षण कराने में ही व्यय हो जाये और उसका सारा उद्देश्य ही समाप्त हो जाये। यदि यह ज्ञान बिलकुल व्यर्थ न भी हो, तब भी, सरकार का काम बल ख़तरा है। यदि सर्वेक्षण का खर्च नियत हो चुका हो तो किस प्रकार कम से कम भ्रातिपूर्ण अनुमान लगाया जा सकता है यह निश्चय करने में सार्विकी के सिद्धान्त हमें मदद पहुँचाते हैं।

उद्योग-धंधों में तो नमूनों के बिना काम ही नहीं चलता। थोक व्यापारी को हजारों की संख्या में भाल लेना पड़ता है। कोई कितना ही अच्छा कारखाना क्यों न

हो उसमें बने हुए माल में थोड़ा बहुत अवश्य ही खराब होता है। यदि एक-एक चीज का निरीक्षण करके उनमें से खराब चीजों को अलग करना हो तो इसके लिए उन्हें एक अलग विभाग कर्मचारियां वा रखना पड़ेगा। इससे उत्पादन का दाम बढ़ जायेगा। यद्यपि थोक व्यापारी को सब माल अच्छा मिलेगा परन्तु इस बढ़े हुए मूल्य के कारण उसे लाभ के बदले हानि ही होगी। किन्तु यदि उसे इस बात का सतोप दिला दिया जाये कि उत्पादन में तो १ प्रतिशत से अधिक माल दोषपूर्ण होने की संभावना बहुत कम है और यदि इस आश्वासन के लिए इतने अधिक निरीक्षण की आवश्यकता न पड़े कि वास्तव में लागत इतनी बढ़ जाये तो संभवतः थोक व्यापारी को सतोप ही जायगा। इस निरीक्षण का किस प्रकार प्रबंध किया जाय कि थोक व्यापारी को भी सतोप हो जाये और खर्च में भी अधिक वृद्धि न हो ? सांख्यिकी के सिद्धान्त इसमें हमें सहायता पहुँचाते हैं।

अभी तो हमने उस दशा में सांख्यिकी के उपयोग का वर्णन किया है जब कि माल बिकने के लिए जाता है। किन्तु उसके पहले भी बहुत-सी समस्याएँ कारखाने वालों के सामने होती हैं। यदि माल खराब तैयार होता है तो उसका कारण खराब कच्चा माल, कल पुर्जों की खराबी या परिचालक की गलती कुछ भी हो सकता है। क्योंकि खराब माल रद्द हो जाता है इसलिए कारखाने को यह पता लगाना बहुत आवश्यक होता है कि खराब माल बनने का क्या कारण है। किस प्रकार के प्रयोग करके इन कारणों का पता लगाया जाये, यह सांख्यिकी का ही काम है। कारण पता चलने पर यदि खराबी कच्चे माल में है तो उसको बदल कर अच्छी सामग्री लेकर खराबी दूर की जा सकती है। यदि कल-पुर्जों में है तो वहाँ खराबी है यह मालूम होने पर इंजीनियर उसे ठीक कर सकते हैं। परिचालक की गलती होने पर उसे उपयुक्त ट्रेनिंग दी जा सकती है या उसे बदला जा सकता है। इन प्रयोगों में जो व्यय होता है वह साधारणतया उस वस्तु के सामने शून्यप्राय ही होता है जो नष्ट हुए पदार्थ के कम होने से होती है। कच्चा माल, मशीन और परिचालक के ठीक होते हुए भी कभी कभी उत्पादन में गड़बड़ी हो जाती है। ऐसी दशा में यदि जरा-जरा-सी खराबी होने पर मशीन की व्यवस्था की जाये तो काम में रुकावट पड़ जाने के कारण व्यय बहुत बढ़ जायेगा। यह भी हो सकता है कि जिस मशीन की व्यवस्था ठीक हो वह भी बिगड़ जाये। इसलिए यह मालूम होना जरूरी है कि क्या वास्तव में ही मशीन में कुछ खराबी है। इसके विपरीत यदि मशीन वास्तव में खराब हो और वह जल्दी ही ठीक न की जाये तो पता नहीं कितना उत्पादन नष्ट हो जाये।

इस द्विविधामयी स्थिति में सांख्यिकी हमारी मदद करती है और नियंत्रण-चार्ट (control chart) की मदद से यह अनुमान लगाया जा सकता है कि मरीजों में व्यवस्था करने की आवश्यकता है या नहीं।

समार में तरह-तरह की बीमारियाँ फैली हुई हैं। इसके साथ ही इन बीमारियों के बारे में सैकड़ों प्रकार की भ्रांतियाँ भी फैली हुई हैं। जितने लोग हैं उतने ही इलाज। बहुत से लोग माने हुए इलाजा की बुराई करते हैं और कहते हैं कि इनको इलाज समझना गलती है। यह एक बिचित्र परिस्थिति है जिसमें यह पता लगाना मुश्किल हो जाता है कि किमका कहना ठीक है और किसका गलत। ऐसी बीमारी कम ही होती है जिनका कोई प्ररोज ठीक ही न हो। बिना इलाज के भी लोग ठीक हो जाते हैं। इस कारण यदि कोई मनुष्य एक विशेष औषधि के लेने के बाद ठीक हो जाता है तो यह कहना उचित नहीं है कि वह बिना औषधि के मर ही जाता। परन्तु कुछ लोग इसको ही औषधि के प्रभावपूर्ण होने का प्रमाण मान लेते हैं। यह पता किस प्रकार लगाया जाय कि कोई औषधि असर कर रही है या नहीं। आप सोचेंगे कि यह एक अजीब समस्या है जिसका हल होना शायद संभव न हो, परन्तु सांख्यिकी के पास इसका भी हल है। यदि कुल रोगियों में से ९० प्रतिशत मर जाते हैं, परन्तु एक विशेष औषधि का सेवन करनेवालों में से केवल १० प्रतिशत मरते हैं तो आप औषधि के प्रभाव को स्वीकार करेंगे अथवा नहीं? आप कह सकते हैं कि यह तो संयोग की बात थी कि इस औषधि का इलाज पाये हुए लोगों में से केवल १० प्रतिशत लोग मरे। सांख्यिकी हमें यह परिकलन करने में सहायता देती है कि केवल संयोगवश इतना अन्तर होना कहीं तक संभव है।

आधुनिक चिकित्सा-विज्ञान (medical science) ने दो दिशाओं में उन्नति की है। एक तो रोग होने के बाद उसके इलाज में और दूसरे बीमारी को फैलाने से रोकने में। इस दूसरी दिशा में प्रगति के लिए यह आवश्यक है कि बीमारी के कारण का पता चलाया जाय। कारण के ज्ञात होने पर उसको दूर करने के उपाय भी मालूम किये जा सकते हैं। जिस प्रकार रोगों के इलाज के बारे में भिन्न-भिन्न धारणाएँ हैं, उसी प्रकार रोगों के कारणों के बारे में भी लोगों में मतभेद है। कोई कहता है कि अमूक रोग मच्छर के काटने से होता है, तो दूसरा बतायेगा कि अमूक वस्तु के खा लेने से यह बीमारी हो जाती है। तीसरा यह कहेगा कि भोजन में अमूक वस्तु की कमी ही इसका कारण है, जब कि चौथा इसे पापों का फल अथवा देवी-देवताओं का प्रकोप समझता है। किसी भी मनुष्य के बीमार होने से पहले यह संभव है कि उसे मच्छर

ने काटा हो, उसने कोई विशेष वस्तु खायी भी हो और उसने भोजन में किसी आवश्यक वस्तु की कमी रही हो। इसी गवाही पर कि उसे मच्छर ने काटा था, यह निश्चय कर लेना कि बीमारी का विशेष कारण यही है, उचित नहीं मालूम होता। इसी प्रकार भोजन के किसी विशेष अंग की कमी की वजह से बीमारी होना अवश्य मभव है, परन्तु किसी विशेष रोगी का अध्ययन करके इसका पता चलाना असभव है। इसके लिए रोगियों के बहुत बड़े समुदाय की जांच करना जरूरी है जिससे यह ज्ञान हो कि उनमें क्या लक्षण समान थे जो उन लोगा में नहीं थे जो रोग से बचे रहे। क्योंकि यहाँ व्यक्ति-विशेष की जांच का नहीं बरन् व्यक्तिगत समुदाय के अध्ययन का प्रश्न है, इसलिए यह साक्षिकी के क्षेत्र में सम्मिलित है। इस प्रकार कारण का पता लगाकर रोगों को फैलने से रोकने में साक्षिकी ने चिकित्सा-विज्ञान की बहुत सहायता की है।

परीक्षा में विद्यार्थियों को बहुधा आपने यह कहते सुना होगा कि भाग्य ने उनका साथ नहीं दिया। जो कुछ उन्होंने नहीं पढ़ा था उसमें से ही प्रश्न रख दिये गये। या अमुक विद्यार्थी बहुत भाग्यशाली है, उसने साल भर कुछ नहीं पढ़ा, परन्तु परीक्षा के पहले दो महीने में उमने जो पढ़ा उसमें से ही सारे प्रश्न आ गये, इसी कारण वह प्रथम श्रेणी में उत्तीर्ण हो गया। आप शायद यह मानेंगे कि ये दावे बिल्कुल बे-बुनियाद नहीं हैं। फिर भी आप यह नहेंगे कि यद्यपि कुछ विद्यार्थियों को, जो योग्यता नहीं रखते, भाग्य से अधिक नंबर मिल सकते हैं तथापि उस विद्यार्थी को—जिसने वास्तव में मेहनत की है और जो योग्य है—कम नंबर नहीं मिल सकते।

लेकिन क्या यह सच है? उत्तर प्रदेश की हाईस्कूल परीक्षा को ही लीजिए। इसमें दो लाख से अधिक विद्यार्थी बैठते हैं। यह असम्भव है कि एक ही परीक्षक इन सबकी कापियाँ जांचे। ये कापियाँ २०० से अधिक परीक्षकों में बाँट दी जाती हैं। क्या दो विद्यार्थी जिन्होंने एक से उत्तर लिखे हैं बराबर नंबर पायेंगे? यदि एक ही उत्तर की दो परीक्षकों द्वारा जांच करवायी जाय तो नंबरों में बहुधा अपेक्षित अंतर पाया जायगा।

इस प्रकार परीक्षाओं में बहुत-सी कमियाँ हैं। इन्हें दूर करने के लिए, विशेष रूप से अमेरिका में, एक नवीन रीति अपनायी गयी है। विद्यार्थी से पाँच या छ लम्बे-लम्बे प्रश्न पूछने के स्थान पर सी या डेढ़ सौ छोटे-छोटे प्रश्न पूछे जाते हैं। इन प्रश्नों से विषय का कोई अंग नहीं बचता। इस प्रकार परीक्षा से भाग्य के प्रभाव का काफी हद तक दूर किया जा सकता है। परीक्षकों के अंतर को दूर करने के लिए भी वहाँ

एक बड़ा सुन्दर नरीका अपनाया जाता है। हर एक प्रश्न के चार या पाँच उत्तर दिये हुए रहते हैं जिनमें केवल एक सही होता है और अन्य सब गलत। परीक्षार्थी को केवल यह बताना होता है कि ठीक उत्तर कौन-सा है। यह पहले से तय ही जाता है कि ठीक होने पर विद्यार्थी को कितने नम्बर मिलेंगे और गलत होने पर कितने नंबर कटेंगे। इस दशा में परीक्षका के अंतर के कारण नवरों में कोई अंतर नहीं पड़ सकता। वास्तव में इस हालत में परीक्षण की कोई आवश्यकता ही नहीं रहती और नवर मशीन द्वारा भी दिये जा सकते हैं।

शायद आपका ध्यान इस ओर गया हो कि परीक्षका के अंतर को दूर करने के लिए जो तरीका अपनाया गया है उसमें फिर भाग्य और मयोग प्रवेश कर गया है। यदि कोई विद्यार्थी केवल अनुमान द्वारा उत्तर का इंगित करे तो भी संयोगवश उसके द्वारा इंगित उत्तर सही हो सकता है। सांख्यिकी इस स्थान पर काम आती है। प्रश्नों की संख्या और उनमें नवर देने का तरीका इस प्रकार का बनाया जाता है कि केवल अनुमान के आधार पर जन्टे नम्बर पाना अगम्य हो जाता है। इसके अतिरिक्त सांख्यिकी का प्रयोग इन सी-डेक सी प्रश्नों के अलग-अलग विश्लेषण में यह जानने के लिए होता है कि कौन-से प्रश्न ऐसे हैं जो अच्छे और बुरे विद्यार्थियों को पहचानने में वास्तव में सहायक हैं। इस प्रकार मानसिक माप को अधिक विद्वसनीय बनाने में सांख्यिकी का काफी भाग है।

पिछले पृष्ठों में आपने उन अनेक क्षेत्रों में से कुछ का परिचय प्राप्त किया है जिनमें सांख्यिकी का एक विशिष्ट स्थान है। आप यह जानने के लिए उत्सुक होंगे कि आन्तरिक सांख्यिकी के ये सिद्धान्त क्या हैं जिनका उपयोगी क्षेत्र इतना विस्तृत है। यह हम पहले ही बना चुके हैं कि सांख्यिकी में जो कार्य सम्मिलित हैं उनमें से एक है प्रेरणा का विश्लेषण करके उन्हें संक्षिप्त रूप में रखना। अगले अध्याय में हम देखेंगे कि आँकड़ों को किस रूप में रखना चाहिए जिसमें हमें उन समुदायों को समझने में सरलता हो जिससे वे गवहित हैं।



अध्याय २

समष्टि और उसका विवरण

§ २.१ समष्टि (population)

इस अध्याय में यह बताया जायगा कि किसी समष्टि के वर्णन के लिए क्या विधि अपनायी जाती है और उसके सांख्यिकीय विवरण में किस प्रकार की विशेषताओं की ओर ध्यान केंद्रित रहता है। व्यवहार में समष्टि का न्यायदर्शन (sample) द्वारा प्रतिनिधित्व किया जा सकता है। परन्तु इस स्थान पर हम प्रतिदर्श और समष्टि में भेद नहीं करेंगे। समष्टि में हमारा तात्पर्य कुछ विशिष्ट इकाइयों के एक समूह से है। हर एक इकाई का कोई गुण (character or attribute) मापा अथवा परखा जा सकता है। ये इकाइयाँ दो प्रकार की हो सकती हैं। प्रथम तो वे जिन्हें साधारण रूप से एक ही समझा जाता है और जिनका अधिक विरलेपण करने पर उनके भागों के गुणों में पूरी इकाई के गुणों से कोई सादृश्य नहीं रहता। इस प्रकार की इकाइयों के उदाहरण हैं मनुष्य, घड़ी और पत्ता। यदि इनके विभिन्न भागों की तुलना की जाय तो आप देखेंगे कि वे एक-दूसरे से इतने भिन्न हैं कि उन्हें सरलता से पृथक्-पृथक् पहचाना जा सकता है। इसके विपरीत कुछ इकाइयाँ इस प्रकार की होती हैं जिनको अपेक्षाकृत छोटी इकाइयों का समूह समझा जा सकता है। इस प्रकार की इकाइयों के उदाहरण हैं सिपाहियों की टुकड़ियाँ, दियासलाइयों का डिब्बा, पुस्तकालय इत्यादि।

§ २.२ चर (variate)

किसी विशेषता के माप को चर (variate or variable) कहते हैं क्योंकि यह विभिन्न इकाइयों के लिए विभिन्न मान (values) धारण कर सकता है। कुछ चर ऐसे होते हैं जिनके लिए दो मानों के बीच का प्रत्येक मान धारण करना संभव है। उदाहरण के लिए मनुष्यों की ऊँचाई इस प्रकार का एक चर है। पाँच

और छ फुट के बीच की सभी ऊँचाइयों के मनुष्य मभव हैं। इस प्रकार के चर को सतत चर (*continuous variable*) कहते हैं। इसके विपरीत परिवार में मनुष्यों की संख्या, पुस्तक में पृष्ठों की संख्या या पुस्तकालय में पुस्तकों की संख्या आदि कुछ ऐसे चर हैं जो कुछ परिमित (*finite*) संस्यक विभिन्न मानों को ही धारण कर सकते हैं। इस प्रकार के चर को असतत चर (*discrete variable*) कहते हैं।

§ २'३ आंकड़ों को सक्षिप्त रूप में रखने की विधि

समष्टि में अनेकों इकाइयाँ होती हैं। यदि उन सबके गुणों के मापों के समूह को आपके सम्मुख रख दिया जाय तो आपको उन्हें समझना और उनमें से तथ्य प्राप्त करना कठिन हो जायगा। निम्नी भी वैज्ञानिक सिद्धान्त के प्रतिपादन के लिए यह नितान्त आवश्यक हो जाता है कि उस ज्ञान को, जो मापों के समूह से प्राप्त होता है, सक्षिप्त रूप में रखा जाय, आवश्यक ज्ञान को अलग किया जाय और अनावश्यक तथा असंगत ज्ञान की उपेक्षा की जाय।

सक्षिप्त करने की सांख्यिकीय विधि में दो विशेष भाग होने हैं —

- (१) आंकड़ों को सारणी अथवा रेखाचित्रों द्वारा सुव्यवस्थित रूप में प्रस्तुत करना,
- (२) कुछ ऐसे सांख्यिकीय मापों का कलन करना जो इन आंकड़ों की विशेषताओं का वर्णन करते हैं।

कुछ उदाहरणों द्वारा इन क्रियाओं को समझने में आसानी होगी। मान लीजिए कि आपके आफिस में २० मनुष्य काम करते हैं। आप इन बीस मनुष्यों के समुदाय का अध्ययन करना चाहते हैं। इस विशेष अध्ययन में आपको जिस चर का विशेष ध्यान है वह है इन मनुष्यों की उम्र। इसके लिए आप प्रत्येक मनुष्य से उसकी उम्र पूछ कर नोट कर लेते हैं। यह उम्र सारणी २१ में दी हुई है।

प्रथम बात जो आपके ध्यान में आयी होगी यह है कि किसी समूह की उम्र सबंधी विशेषताओं के वर्णन में उस समूह के मनुष्यों के नामों का कोई स्थान नहीं है। इस प्रकार के असंगत ज्ञान की उपेक्षा की जा सकती है। इसके अतिरिक्त इन उम्रों को विशेष क्रम में रखने पर उसके समझने में सहायता मिल सकती है। ऊपर की सारणी के संगत भाग को हम निम्नलिखित सक्षिप्त रूप में रख सकते हैं।

सारणी सख्या 21

आफिस के मनुष्या के नाम और उनकी उम्र

क्रम सख्या	नाम	उम्र निकटतम वर्षों में	क्रम सख्या	नाम	उम्र निकटतम वर्षों में
1	2	3	1	2	3
1	अपोष्या सिंह	25	11	विमल चंद्र	25
2	अवध बिहारी	23	12	नवीन	25
3	कमल कृष्ण	28	13	बलवत राम	28
4	नरसिंह	28	14	बाल कृष्ण	25
5	राज्य प्रकाश	26	15	निमल	27
6	ओम प्रकाश	27	16	हरी प्रसाद	27
7	हुकुम चंद्र	25	17	काशिम	28
8	याकूब	27	18	जय प्रकाश	25
9	रमेश चंद्र	26	19	केवल राम	25
10	रमेश प्रसाद	28	20	अनोखे लाल	25

सारणी सख्या 22

आफिस के मनुष्यों की उम्र का वितरण

क्रम सख्या	उम्र निकटतम वर्षों में	वारवारता
i	x_i	f_i
(1)	(2)	(3)
1	23	1
2	24	0
3	25	8
4	26	2
5	27	4
6	28	5
कुल		20

इससे हमें यह पता चलता है कि भिन्न-भिन्न अवस्था के कितने मनुष्य इस समुदाय में हैं। बारबारता (frequency) के अर्थ हैं उन इकाइयों की संख्या जिनमें माप समान है। उदाहरणार्थ 25 वर्ष की उम्र के मनुष्यों की बारबारता इस समुदाय में 8 है। इस प्रकार की सारणी को बारबारता सारणी (frequency table) कहते हैं। इसके द्वारा समस्त माप के बारबारता-वटन अथवा वितरण (frequency distribution) का पता चल जाता है।

यदि हम यह जानना चाहे कि 27 वर्ष अथवा उससे कम अवस्था के कितने मनुष्य आपके आफिस में हैं तो हमें उन सब बारबारताओं का योग करना होगा जो 27 वर्ष और उससे कम उम्र के मनुष्यों की हैं। इस आफिस में यह संचयी बारबारता (cumulative frequency) $1+0+8+2+4=15$ है। इस प्रकार ऊपर की हुई बारबारता सारणी को सहायता से एक संचयी बारबारता सारणी बनायी जा सकती है।

सारणी संख्या 23

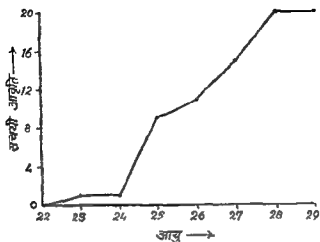
आम के आफिस के मनुष्यों की उम्र की संचयी बारबारता सारणी

क्रम-संख्या i	उम्र निकटतम वर्षों में x_i	संचयी बारबारता F_i
(1)	(2)	(3)
1.	23	1
2	24	1
3.	25	9
4	26	11
5	27	15
6	28	20

§ 2.4 आँकड़ों का रेखाचित्रों द्वारा निरूपण

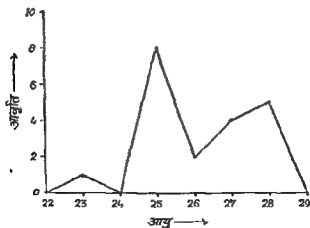
ये संचयी बारबारताएँ एक ग्राफ पर बिन्दुओं द्वारा निरूपित की जा सकती हैं। इन बिन्दुओं को मिलाती हुई जो रेखा खींची जाती है उसे संचयी बारबारता का रेखाचित्र (cumulative frequency diagram) अथवा तोरण (ogive) कहते हैं।

इसी प्रकार बारबारताओं को ग्राफ पर बिन्दुओं द्वारा निरूपित करने और क्रमगत बिन्दुओं को रेखाओं द्वारा मिला देने पर बारबारता का रेखाचित्र बन जाता



चित्र १—संख्या बारंबारता

है। उस टेडी-मेडी रेखा को जो इन बिंदुओं को मिलाती है, बारंबारता-बहुभुज (frequency polygon) कहते हैं।

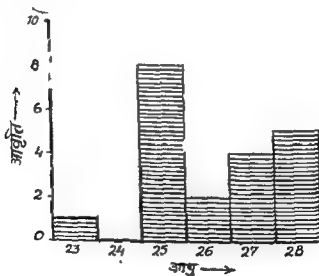


चित्र २—आवृत्ति बहुभुज

यदि चर कुछ परिमित (finite) मानों को ही धारण कर सकता है तो

ग्राफ में इन मानों के लिए बारवारता को विदुआ द्वारा सूचित किया जा सकता है। यदि इन विदुओं में भुजाक्ष (axis of abscissa) पर ऊर्ध्व रेखाएँ खींची जायें तो उनकी लंबाई में इन बारवारताओं का अधिक स्पष्ट आभास हो जाता है। इस प्रकार निरूपण को दण्ड चित्र (bar diagram) कहते हैं।

इसके विपरीत यदि चर सतत हो तो चर के पराम (range) को कुछ भागों में विभाजित कर दिया जाता है। सारणी में प्रत्येक भाग के लिए चर की बारवारता दी जाती है। ग्राफ में इन भागों को भुजाक्ष पर अंतरालों से सूचित किया जाता है। प्रत्येक अंतराल पर ऐसा समकोण चतुर्भुज बनाया जा सकता है जिसका क्षेत्रफल उस अंतराल में चर की बारवारता को सूचित करना हो। बारवारता के इस प्रकार के निरूपण को आयत-चित्र (histogram) कहते हैं।



चित्र ३—आयत चित्र

आयत चित्र अथवा बारवारता बहुभुज दोनों से हमें बारवारता सारणी में दी हुई सब सूचना प्राप्त हो जाती है। बहुधा चित्र द्वारा वे विशेषताएँ स्पष्ट ही जाती हैं जिनको अक्षों के रूप में समझना अपेक्षाकृत कठिन है। इसी प्रकार सचयी बारवारता चित्र द्वारा सचयी बारवारता की विशेषताएँ अधिक स्पष्ट हो जाती हैं।

§ २५ चर के परास का विभाजन

एक बात पर शायद आपका ध्यान गया होगा। उम्र एक सतत चर है। जिन मनुष्यों की उम्र २५ वर्ष लियी हुई है वास्तव में उन सबकी उम्र एकदम समान नहीं है। उनमें महीने अथवा दिनों का अंतर हो सकता है। ऐसी दशा में भाग के हर सूक्ष्म-तम भाग के लिए बारबारता-चित्र बनाना नितान्त अयमभव है। इसलिए इसके ध्यान पर उम्र के परास (range) को कुछ भागों में विभाजित कर लिया जाता है और केवल उन्ही भागों के लिए बारबारता-सारणी बनायी जाती है। उदाहरण के लिए ऊपर की सारणी में २३ वर्ष का अर्थ है २२ ५ से लेकर २३ ५ वर्ष तक का अंतराल। आमतौर पर इसको ही ध्यान में रखकर बनाया जाता है।

यदि चर परिमित हो तो भी परास को इस प्रकार विभाजित करने की आवश्यकता पड़ सकती है। यह तब होता है जब छोटी इकाइयों की तुलना में परास बहुत अधिक हो। उदाहरणार्थ यदि एक नगर के मनुष्यों की आय के अनुसार बारबारता-सारणी बनायी जाय तो आयों का परास दैन्य से लेकर दस हजार रुपये मासिक तक हो सकता है। यदि एक एक रुपये की आय के अंतर से बारबारता मालूम की जाय तो न केवल बहुत अधिक मेहनत पड़ेगी वरन् इस बृहद् सारणी को समझना और उससे किसी तत्त्व को प्राप्त करना अयमभव हो जायगा। इसलिए परास को अपेक्षाकृत कम भागों में विभाजित करना आवश्यक हो जाता है। साधारणतया बीस या पच्चीस से अधिक भागों में विभाजित करने से सारणी को समझने में कठिनाई पड़ती है।

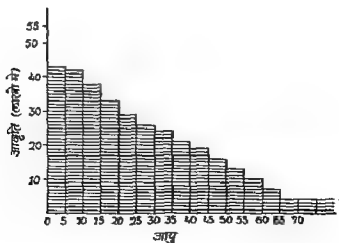
यदि हो सके तो इन भागों का—जिनमें परास को विभाजित किया जाता है—बराबर होना अच्छा रहता है। परन्तु कई बार भागों के बराबर होने से कठिनाई हो जाती है। उदाहरण के लिए आयों के परास को यदि बीस भागों में बाँटा जाय तो प्रत्येक भाग पाँच सौ रुपयों का प्रतिनिधित्व करेगा। इनमें से केवल दो भाग १,००० से कम आय का प्रतिनिधित्व करेंगे। और अठारह भाग एक हजार से लेकर दस हजार रुपये तक की आय का। नगर की एक लाख से अधिक जनसंख्या में शायद आठ दस मनुष्य ही ऐसे होंगे जिनकी मासिक आय एक हजार रुपये से अधिक हो। यह स्पष्ट है कि आयों के ऊपर लिखित बराबर विभाजन द्वारा हम बहुत सा ज्ञान खो देंगे। इस प्रकार की स्थिति में पहिले छोटे और फिर क्रमशः बड़े भागों में परास को विभाजित करना आवश्यक हो जाता है।

नीचे बारबारता-सारणी और उसके ग्रेफाचित्रीय निरूपण (graphic representation) के कुछ उदाहरण दिये हुए हैं ।

सारणी सरया 24

उत्तर प्रदेश के पुरुषों की उम्र-बारबारता-सारणी

क्रम संख्या	उम्र का अंतराल (वर्षों में)	पुरुष-संख्या (सैकड़ों में)	क्रम संख्या	उम्र का अंतराल (वर्षों में)	पुरुष-संख्या (सैकड़ों में)
1	(2)	(3)	(4)	(5)	(6)
1	0—5)	42 694	9	40—45)	18 516
2	5—10)	41 965	10	45—50)	15 934
3	10—15)	37 671	11	50—55)	12 967
4	15—20)	33 008	12	55—60)	9 870
5	20—25)	29 112	13	60—65)	6 876
6	25—30)	26 296	14	65—70)	4 349
7	30—35)	23 793	15	70—	6 736
8	35—40)	21 302		कुल	330 989

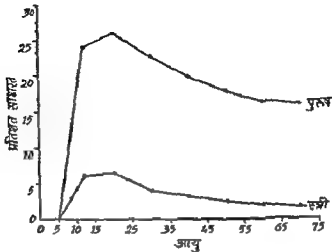


चित्र ४—उत्तर प्रदेश के पुरुषों की आयु-आकृति का आयत चित्र

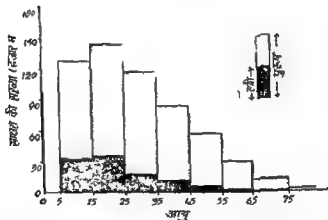
सारणी संख्या 2.5

उत्तर प्रदेश में उम्र और साक्षरता—(मर्यादाएँ इस की इलाहियों में)

	[0-5]		[15-25]		[25-35]		[35-45]		[45-55]		[55-65]		[65-75]		[75-]	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
पु																
१	साक्षर	0	29,574	99,254	143,441	116,853	80,350	55,550	28,738	11,260	4,357					
२	कुल	426,063	419,039	418,368	552,691	507,412	406,482	307,213	174,655	70,197	28,582					
३	प्रतिशत-साक्षर	0.00	7.06	23.72	25.95	23.03	19.77	18.08	16.45	16.04	15.24					
स्त्रि																
४	साक्षर	0	9,777	22,107	33,546	18,700	10,626	6,408	3,672	1,284	481					
५	कुल	415,794	383,741	351,682	510,778	449,748	343,205	254,988	151,069	69,858	33,029					
बाँ																
६	प्रतिशत-साक्षर	0.00	2.55	6.29	6.57	4.16	3.10	2.51	2.43	1.84	1.46					



चित्र ५—उत्तर प्रदेश में प्रतिशत साक्षरता



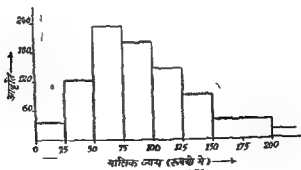
चित्र ६—उ० प्र० में साक्षरता का आयत चित्र

नोट—अंतराल $[a, b)$ से उन सब सख्याओं के समुदाय को सूचित किया जाता है जो b से छोटी और a के बराबर अथवा a से बड़ी है। इसी प्रकार $(a, b]$ से उन सख्याओं के समुदाय को सूचित किया जाता है जो a से बड़ी और b के बराबर अथवा b से छोटी है।

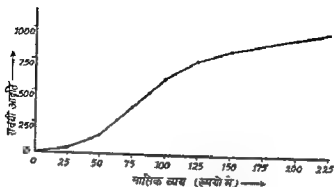
सारणी सख्या 26

फरीदाबाद के एक हजार परिवारों का प्रतिमास-व्यय के अनुसार वितरण

क्रम	प्रतिमास व्यय (रुपयों में)	परिवारों की सख्या	सन्धियों बारवारता
(1)	(2)	(3)	(4)
1	[0—25 5)	34	34
2	[25 5—50 5)	123	156
3	[50 5—75 5)	234	390
4	[75 5—100 5)	202	592
5	[100 5—125 5)	146	738
6	[125 5—150 5)	94	832
7	[150 5—200 5)	100	932
8	[200 5—	68	1,000



चित्र ७—फरीदाबाद के परिवारों का मासिक व्यय के अनुसार वितरण-भायत चित्र



चित्र C—फरीदाबाद के परिवारों का मासिक व्यय के अनुसार संचयी आवृत्ति चित्र

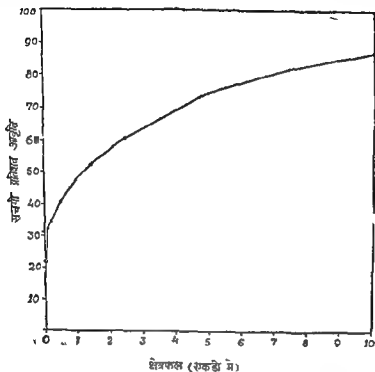
सारणी सख्या 27

अधिकृत जमीन के क्षेत्रफल के अनुसार भारतीय ग्राम परिवारों का प्रतिशतता वितरण

अधिकृत क्षेत्रफल (एकड़ों में)	परिवारों की प्रतिशतता	अधिकृत क्षेत्रफल (एकड़ों में)	परिवारों की प्रतिशतता
(1)	(2)	(1)	(2)
[0—0 005)	22 00	[7 495—9 995)	04 71
[0 005—0 045)	09 78	[9 995—14 995)	05 12
[0 045—0 095)	02 74	[14 995—19 995)	02 66
[0 095—0 495)	06 12	[19 995—24 995)	01 43
[0 495—0 995)	06 25	[24 995—29 995)	01 07
[0 995—1 495)	05 29	[29 995—39 995)	01 07
[1 495—2 495)	08 58	[39 995—49 995)	00 50
[2 495—4 995)	13 66	[49 995—74 995)	00 55
[4 995—7 495)	08 16	[74 995—	00 31

ऊपर के बारंबारता चित्र और आयत चित्रों को देखकर एक बात आपके ध्यान में आयी होगी। प्रायः सभी आकड़ों में एक केंद्रीय प्रवृत्ति (central tendency) है। किसी विशेष भाग में बारंबारता अधिकतम है और उसके दोनों ओर बारंबारता क्रमशः कम होता चली जाती है। बहुत छोटी अथवा बहुत

बड़ी राशियों की बारबारताएँ कम हैं। यदि इस केन्द्रीय प्रवृत्ति का और इसके दोनों ओर की बारबारताओं के प्रसार (dispersion) का भी हमें कोई माप



चित्र ९—भारतीय ग्राम परिवारों का अधिकृत क्षेत्रफल के अनुसार वितरण—सबकी आवृत्ति चित्र का एक भाग

(measure) मिल जाय तो मोटे रूप में हमें समष्टि के स्वरूप का ज्ञान ही जाता है। नीचे केन्द्रीय प्रवृत्ति के कुछ मापों की व्याख्या दी हुई है।

§ २६ केन्द्रीय प्रवृत्ति के कुछ माप

(क) समान्तर माध्य (arithmetic mean) या केवल माध्य (mean)

यदि समष्टि की सब इकाइयों के चरों के मानों को जोड़कर उसमें इकाइयों की कुल संख्या का भाग लगाया जाय तो फल को समानान्तर माध्य अथवा केवल माध्य

कहते हैं। यदि $x_1, x_2, x_3, \dots, x_n$ चरों के मान हैं तो माध्य—जिसे साधारणतया \bar{x} से सूचित किया जाता है—को निम्न लिखित सूत्र द्वारा प्राप्त किया जा सकता है।

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad \dots (2.1)$$

मानों के योग को सूत्र रूप में लिखने की एक और उत्तम विधि है। $x_1 + x_2 + x_3 + \dots + x_n$ लिखने के स्थान में हम इस योग को संक्षिप्त रूप में $\sum_{i=1}^n x_i$ लिख सकते हैं।

उदाहरण के लिए $\sum_{i=1}^4 x_i$ का अर्थ है $x_1 + x_2 + x_3 + x_4$ ।

यदि आँकड़े बारंबारता सारणी के रूप में दे रखे हों तो माध्य प्राप्त करने के लिए निम्नलिखित सूत्र का प्रयोग किया जा सकता है

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{\sum_{i=1}^k f_i} \quad \dots (2.2)$$

जहाँ कुल k अंतरालों में परास को विभाजित किया गया हो और i वें अंतराल का मध्य बिन्दु x_i तथा इस अंतराल में बारंबारता f_i हो। शर्द्धि एक अंतराल में भी सब मान उसके मध्य बिन्दु के बराबर नहीं होते फिर भी यदि अंतराल बहुत बड़ा न हो तो इन सब मानों के माध्य को अंतराल का मध्य बिन्दु मान लेने से कोई विशेष हानि नहीं होती।

आइए हम इस माप से परिचय प्राप्त करने के लिए पूर्व परिचित बारंबारता सारणियों की सहायता लें।

(१) सारणी सख्या 2.2—आफिस में काम करने वाले मनुष्यों की औसत उम्र क्या ? यदि इस औसत को \bar{x} से सूचित किया जाय तो—

$$\bar{x} = \frac{\sum_{i=1}^{20} x_i f_i}{\sum_{i=1}^{20} f_i}$$

$$\begin{aligned}
 &= \frac{(23 \times 1) + (24 \times 0) + (25 \times 8) + (26 \times 2) + (27 \times 4) + (28 \times 5)}{1 + 0 + 8 + 2 + 4 + 5} \text{ वर्ष} \\
 &= \frac{23 + 0 + 200 + 52 + 108 + 140}{20} \text{ वर्ष} \\
 &= \frac{523}{20} \text{ वर्ष} \\
 &= 26.15 \text{ वर्ष}
 \end{aligned}$$

यदि सारणी में अंतराल बराबर हो, जैसा कि ऊपर के उदाहरण में है, तो माध्य का परिकलन बहुत सरल हो जाता है। इस अंतराल को इकाई मानकर और किसी भी स्वेच्छ (arbitrary) मूलबिंदु (origin) को लेकर अंतरालों के मध्य बिंदुओं को नवीन सख्याओं के द्वारा निरूपित किया जा सकता है। इस प्रकार नीचे दी हुई सारणी प्राप्त होगी।

सारणी सख्या 2 2:2

क्रम सख्या i	मध्य बिंदु (वर्षों की इकाई में) x_i	25 वर्ष को मूलबिंदु और 1 वर्ष को इकाई मानकर मध्यबिंदु का निरूपण $(i-3) = m_i$	बारबारता f_i
(1)	(2)	(3)	(4)
1	23	-2	1
2	24	-1	0
3	25	0	8
4	26	1	2
5	27	2	4
6	28	3	5

ऊपर दिये हुए विन्यास (arrangement) से यह स्पष्ट है कि किसी भी अंतराल के मध्यबिंदु का पूर्व-निरूपित मान $x_i = 25 + m_i \times 1$ वर्ष

$$\begin{aligned}
 \therefore \bar{x} &= \frac{\sum_{i=1}^6 x_i f_i}{\sum_{i=1}^6 f_i} \\
 &= \frac{\sum_{i=1}^6 \{25 + m_i\} f_i}{\sum_{i=1}^6 f_i} \text{ वर्ष}
 \end{aligned}$$

$$= 25 + \frac{\sum_{i=1}^n m_i f_i}{\sum_{i=1}^n f_i} \text{ वर्ष}$$

$$= 25 + \bar{m}$$

जहाँ \bar{m} मध्यविंदुआ के नवीन माना का माध्य है।

$$= 25 + \frac{(-2 \times 1) + (1 \times 2) + (2 \times 4) + (3 \times 5)}{20}$$

$$= 25 + \frac{23}{20} \text{ वर्ष}$$

$$= 26.15 \text{ वर्ष}$$

इस उदाहरण में नवीन और आरंभिक मध्यविंदुओं के अंतराल समान थे। इसलिए अब हम एक दूसरा उदाहरण लेंगे जिसमें ये अंतराल बराबर न हों। सारणी संख्या 24 इसके लिए उपयुक्त होगी। यहाँ हम केवल प्रथम 14 अंतरालों पर विचार करेंगे। भ्रम लीजिए आरंभ से अंतराल h ही और नवीन मध्यविंदुओं के लिए x_k को मूलविंदु माना गया हो तो—

$$x_1 = x_k + (1 - k) h$$

$$= x_k + m_i h$$

$$\therefore \bar{x} = \frac{\sum (x_k + m_i h) f_i}{\sum f_i}$$

$$= x_k + \bar{m} h \quad (23)$$

सारणी संख्या 242

क्रम संख्या	आरंभिक मध्यविंदु	नवीन मध्यविंदु	बारबारता	क्रम संख्या	आरंभिक मध्यविंदु	नवीन मध्यविंदु	बारबारता
1	x_1	m_1	f_1	1	x_1	m_1	f_1
(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
1	25	-6	42 694	8	37.5	1	21,202
2	7.5	-5	41 965	9	42.5	2	18,516
3	12.5	-4	37 671	10	47.5	3	15,934
4	17.5	-3	33 008	11	52.5	4	12,967
5	22.5	-2	29 112	12	57.5	5	9,870
6	27.5	-1	26 296	13	62.5	6	6 876
7	32.5	0	23 793	14	67.5	7	4,349

उत्तर प्रदेश के पुरुषों की माध्य आय $\bar{x} = (32\ 5 + \bar{m} \times 5)$ वर्ष

$$\bar{m} = [1 \times (21,202 - 26,296) + 2 \times (18\ 516 - 29,112)$$

$$+ 3 \times (15,934 - 33,008) + 4 \times (12\ 967 - 37\ 671)$$

$$+ 5 \times (9,870 - 41,965) + 6 \times (6,876 - 426,94)$$

$$+ 7 \times 4,349] \times \frac{1}{331,989}$$

$$= \frac{-1}{331,989} [5\ 094 + 2 \times 10\ 596 + 3 \times 17,074$$

$$+ 4 \times 24\ 704 + 5 \times 32\ 095 + 6 \times 35,818 - 7 \times 4,349]$$

$$= \frac{-521,344}{331,989}$$

$$= -1\ 57$$

$$\therefore \bar{x} = (32\ 50 - 1\ 57 \times 5) \text{ वर्ष}$$

$$= (32\ 50 - 7\ 85) \text{ वर्ष}$$

$$= 24\ 65 \text{ वर्ष}$$

(ख) केंद्रीय प्रवृत्ति का एक अन्य माप माध्यिका (median) है। जब सब प्रेक्षणों को उनके मानों के बढ़ते हुए परिमाणों के अनुसार विन्यास किया जाता है तो मध्य के प्रेक्षण को माध्यिका कहते हैं। यदि इस विन्यास के अनुसार प्रथम प्रेक्षण का मान x_1 , द्वितीय या x_2 , अन्तिम का x_{2m+1} हो तो माध्यिका x_{m+1} है। यदि कुल प्रेक्षणों की संख्या विषम (odd) न होकर सम (even)— $2m$ हो तो माध्यिका मध्य के दो मानों x_m और x_{m+1} का माध्य $\frac{1}{2}(x_m + x_{m+1})$ होती है।

यदि आँकड़े बारबारता सारणी के रूप में दिये गये हों तो कुछ अधिक परिकलन की आवश्यकता पड़ती है। सचयी बारबारता के आधार पर हम यह आसानी से मालूम कर सकते हैं कि माध्यिका कौन से अंतराल में स्थित है। इस अंतराल को माध्यिका अन्तराल (median interval) कहते हैं। मान लीजिए कुल प्रेक्षणों की संख्या n है। सचयी बारबारताएँ क्रमशः $F_1, F_2, F_3, \dots, F_k, \dots, F_n$ हैं जहाँ कुल अंतरालों की संख्या s है। यदि $F_k < \frac{n}{2} \leq F_{k+1}$ तो माध्यिका अंतराल $(k+1)$ वाँ है। मान लीजिए अन्तरालों के सीमान्त बिंदु क्रमशः $x_1, x_2,$

. x_k है। इस परिवर्तन के लिए यदि यह मान लिया जाय कि अन्तराल में किसी भाग में बारबारता उस भाग की लंबाई की समानुपाती (proportional) है तो

$$\text{माध्यिका} = x_k + (x_{k+1} - x_k) \times \frac{\left(\frac{n}{2} - F_k\right)}{(F_{k+1} - F_k)} \quad \dots (24)$$

उदाहरण

(१) सारणी सख्या २३ में $n=20$ तीसरे अन्तराल तक संचित आवृत्ति ९, तथा चौथे तक ११ है। इसलिए माध्यिका अन्तराल चौथा है। इस अन्तराल का प्रथम बिंदु २५.५ वर्ष है तथा अन्तिम बिंदु २६.५ वर्ष है।

$$\cdot x_k = 25.5 \text{ वर्ष}$$

$$x_{k+1} = 26.5 \text{ वर्ष}$$

$$\frac{n}{2} = 10$$

$$F_k = 9$$

$$F_{k+1} = 11$$

$$\cdot \text{माध्यिका} = 25.5 + 1 \times \frac{1}{2} \text{ वर्ष}$$

$$= 26 \text{ वर्ष}$$

(२) सारणी सख्या २६ में

$$x_k = 75.50 \text{ रुपये}$$

$$x_{k+1} = 100.50 \text{ रुपये}$$

$$\frac{n}{2} = 500$$

$$F_k = 390$$

$$F_{k+1} = 592$$

$$\cdot \text{माध्यिका} = 75.50 + 25 \times \frac{110}{202} \text{ रुपये}$$

$$= 75.50 + 13.61 \text{ रुपये}$$

$$= 89.11 \text{ रुपये}$$

(ग) बहुलक (mode) केन्द्रीय प्रवृत्ति का तीसरा माप है। यह चर का वह मान है जिसकी बारबारता सबसे अधिक होती है। यदि आंकड़े बारबारता

सारणी के रूप में दिये हुए हों तो उस अंतराल को जिसमें बारबारता सबसे अधिक होती है बहुलक-अंतराल (*modal interval*) कहते हैं। बहुलक के विरोध मान के लिए उस अंतराल का मध्य बिंदु लिया जाता है जिसमें बारबारता सबसे अधिक हो।

उदाहरण —

(१) सारणी सख्या 2.2 में सबसे अधिक बारबारता 8 उस अंतराल में है जिसका मध्यबिंदु 25 वर्ष है। इसलिए आफिस में आयु का बहुलक 25 वर्ष है।

(२) सारणी सख्या 2.4 में सबसे अधिक बारबारता प्रथम अंतराल में है जिसका मध्यबिंदु 2.5 वर्ष है। इसलिए उत्तर प्रदेश के पुरुषों की आयु का बहुलक 2.5 वर्ष है।

(३) सारणी सख्या 2.5 के दो भाग हैं एक में पुरुषों के लिए और दूसरे में स्त्रियों के लिए साक्षरों की बारबारताएँ उन्न के अनुसार दी गयी हैं। इनमें बहुलक का परिक्लन करने के लिए हमें दूसरी विधि अपनानी पड़ेगी क्योंकि सब अंतराल समान नहीं हैं। यह स्पष्ट है कि यदि किसी अंतराल को दूसरी की अपेक्षा बहुत बड़ा बना दिया जाय तो उसमें बारबारता अपेक्षाकृत अधिक होगी। हम चाहेंगे कि हमारा माप जहाँ तक हो सके उस विधि से स्वतन्त्र हो जिसके अनुसार कुल परास को अंतराल में विभाजित किया जाना है। इसके लिए युक्तिसंगत यह है कि अंतराल की प्रति इकाई के लिए बारबारता जिस अंतराल में अधिक हो उसे बहुलक-अंतराल समझा जाय और बहुलक को उसका मध्य बिंदु माना जाय। उदाहरण के लिए सारणी सख्या 2.5 में साक्षर पुरुषों की प्रति इकाई बारबारता अंतराल [10 - 15) में $\frac{99,254}{5} = 19,850.8$ है जो अन्य अंतरालों की प्रति इकाई बारबारता से अधिक है। अंतराल (15 - 25) में यह प्रति-इकाई बारबारता केवल $\frac{143,441}{10} = 14,344.1$ है। इस प्रकार वास्तविक बहुलक और सारणी से प्राप्त बहुलक में अंतर कम हो जाता है। सारणी सख्या 2.5 में, इस दृष्टिकोण से, स्त्री व पुरुषों दोनों के लिए बहुलक 12.5 वर्ष है। यानी साक्षर लोगों में सबसे अधिक सख्या 12 से 13 वर्ष तक के व्यक्तियों की है।

§ २७ प्रसार के कुछ माप

केन्द्रीय प्रवृत्ति के इन तीन मापों के आधार पर हमें समष्टि का कुछ ज्ञान प्राप्त होता है। परन्तु यह यथेष्ट नहीं है। आपने यह कहावत सुन ही रखी होगी कि

“लेखा जोसा जमो वा त्यो, सारा कुनवा डूबा क्या ?” एवं मनुष्य परिवार सहित किमी नदी को पार कर रहा था। जब उसे मालूम हुआ कि नदी में पानी की औसत गहराई केवल एक फुट है तो नाव चलाना असंभव समझकर और उसका ध्वंस बचाने के लिए उसने पैदल ही नदी पार करने का फैसला किया। परन्तु बीच में नदी की गहराई थोम फुट तक थी और मारा कुनवा पैदल नदी पार करने के प्रयत्न में डूब गया। यह स्पष्ट है कि इन केन्द्रीय प्रवृत्ति के मापों के दोनों ओर बारंबारताओं के प्रसार (dispersion) को समझने के लिए कुछ अन्य मापों की भी आवश्यकता है। इनमें से कुछ मुख्य माप नीचे दिये हुए हैं।

(क) परास (range) चर के महत्तम और न्यूनतम मानों के अंतर को कहते हैं। उदाहरण के लिए सारणी मख्या 22 में न्यूनतम आयु 22.5 वर्ष और महत्तम 28.5 वर्ष है। इसलिए आफिस में काम करने वाली की आयु का परास 6 वर्ष है।

(ख) मानक विचलन (standard deviation) चर के किसी विशेष मान x_i का माध्य \bar{x} से विचलन (deviation) $(x_i - \bar{x})$ है। कुल विचलनों का योग शून्य है।

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - n\bar{x}$$

$$= 0$$

$$\text{क्योंकि } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

परन्तु इन विचलनों का वर्ग माध्य $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ शून्य नहीं है

क्योंकि इस योग में प्रत्येक पद धनात्मक है। इस वर्ग माध्य का वर्गमूल (square root) प्रसार का एक अन्य उपयुक्त माप है। इसको विचलन-वर्ग माध्य-मूल (root mean square deviation) या साधारणतः मानक विचलन कहते हैं। लघुरूप में हम इसको मा० वि० से सूचित करेंगे।

$$\therefore (\text{मा० वि०})^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (25)$$

यदि आंकड़े बारबारता सारणी के रूप में दिये हुए हों तो—

$$(\text{मा० वि०})^2 = \frac{\sum_{i=1}^k f_i (x_i - \bar{x})^2}{\sum_{i=1}^k f_i} \quad \dots \quad (2.6)$$

जहाँ सारणी में कुल k अंतराल हैं और 1 वें अंतराल में बारबारता f_1 है। यह तो हमें सूत्र (2.2) द्वारा पता ही है कि—

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{\sum_{i=1}^k f_i}$$

सख्यात्मक अभिगणना (arithmetical computations) के लिए सूत्र (2.5) और सूत्र (2.6) में वगं-योग को अधिक सुविधाजनक रूप में रखा जा सकता है।

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i^2 - 2x_i \bar{x} + \bar{x}^2) \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - n\bar{x}^2 \quad \dots \quad (2.7) \end{aligned}$$

$$\begin{aligned} \sum_{i=1}^k f_i (x_i - \bar{x})^2 &= \sum_{i=1}^k f_i (x_i^2 - 2x_i \bar{x} + \bar{x}^2) \\ &= \sum_{i=1}^k f_i x_i^2 - 2\bar{x} \sum_{i=1}^k f_i x_i + \bar{x}^2 \sum_{i=1}^k f_i \\ &= \sum_{i=1}^k f_i x_i^2 - \left(\sum_{i=1}^k f_i \right) \bar{x}^2 \quad \dots \quad (2.8) \end{aligned}$$

$$\therefore (\text{मा० वि०})^2 = \frac{\sum_{i=1}^k f_i x_i^2}{\sum_{i=1}^k f_i} - \bar{x}^2 \quad \dots \dots (2'6'2)$$

उदाहरण—

(१) सारणी सख्या (2'2) $\bar{x} = 26.15$ वर्ष

$$\begin{aligned} (\text{मा० वि०})^2 &= \left[\frac{\{(23)^2 \times 1\} + \{(25)^2 \times 8\} + \{(26)^2 \times 2\} + \{(27)^2 \times 4\}}{20} \right. \\ &\quad \left. + \{(28)^2 \times 5\} - (26.15)^2 \right] (\text{वर्ष})^2 \\ &= \left[\frac{13,717}{20} - (26.15)^2 \right] (\text{वर्ष})^2 \\ &= [685.8500 - 683.8225] (\text{वर्ष})^2 \\ &= 2.0275 (\text{वर्ष})^2 \end{aligned}$$

ऊपर हमें 23 से लेकर 28 तक के अंकों के वर्गों का परिकलन करना पडा। यदि मान और बड़े बड़े होते तो यह परिकलन काफी कठिन हो जाता। हम देख चुके हैं कि माध्य का परिकलन स्वेच्छ मूल बिंदु को लेने से बहुत सरल हो जाता है। मानक विचलन का वर्ग भी तो एक माध्य है। इसलिए इसके परिकलन को भी स्वेच्छ मूल बिंदु लेकर सरल बनाया जा सकता है।

यदि मान a को स्वेच्छ मूल बिंदु माना जाये और

$$\begin{aligned} \text{तो} \quad x_i &= a + x'_i \\ \bar{x} &= a + \bar{x}' \end{aligned}$$

$$\text{जहाँ} \quad \bar{x}' = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\text{और} \quad \bar{x}' = \frac{1}{n} \sum_{i=1}^n x'_i$$

$$\therefore \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n \{(a + x'_i) - (a + \bar{x}')\}^2$$

$$= \sum_{i=1}^n (x_i' - \bar{x}')^2$$

$$= \sum_{i=1}^n x_i'^2 - n \bar{x}'^2 \quad \dots \quad (29)$$

यदि आंकड़े ऐसी बारबारता सारणी के रूप में दिये हुए हों जिसमें अंतराल बराबर हों, तो सक्षारमक परिकलन को निम्नलिखित विधि से सरल बनाया जा सकता है।

$$x_i = x_r + (i-r)h$$

$$= x_r + m_i h$$

जहाँ x वें अंतराल के मध्य बिंदु x_r को स्वेच्छ मूल-बिंदु मान लिया गया हो और अंतराल का मान h हो।

$$\therefore x_i - \bar{x} = (m_i - \bar{m}) h$$

$$\text{जहाँ } m = \frac{\sum_{i=1}^k m_i f_i}{\sum_{i=1}^k f_i}$$

$$\therefore \sum_{i=1}^k f_i (x_i - \bar{x})^2 = h^2 \sum_{i=1}^k f_i (m_i - \bar{m})^2$$

$$= h^2 \left(\sum_{i=1}^k f_i \right) \times (m_i \text{ का मा० वि०})^2$$

..... (210)

आइए, हम ऊपर के उदाहरण में मा० वि० का परिकलन इस सुगम रीति से करें। पहिले की भाँति 25 वर्ष की स्वेच्छ मूल-बिंदु मान लीजिए अर्थात् $r=3$ तथा $h=1$ है। अतः $x_i = 25 + (i-3)$ ।

$$20 \times (\text{मा० वि०})^2 = \{(-2)^2 \times 1 + (1^2 \times 2) + (2^2 \times 4) + (3^2 \times 5)\}$$

$$= 20 \times (1.15)^2 \quad (\text{वर्ष})^2$$

$$= [67 - 26.45] \quad (\text{वर्ष})^2$$

$$\therefore (\text{मा० वि०})^2 = \left[\frac{40.55}{20} \right] \quad (\text{वर्ष})^2$$

$$= 2.0275 \quad (\text{वर्ष})^2$$

मानक विचलन के परिकलन के पूर्व उसके वर्ग का परिकलन करना पड़ता है। इस वर्ग को प्रसरण (*variance*) कहते हैं।

(ग) माध्य-विचलन (*mean deviation*)—प्रसार के माप के लिए भिन्न भिन्न विचलनों ($x_i - \bar{x}$) के योग से काम नहीं चल सकता क्योंकि इसका मान प्रत्येक समष्टि के लिए शून्य होता है। परन्तु यदि विचलनों के निरपेक्ष मानों (*absolute values*) अर्थात् धन अथवा ऋण चिह्न विहीन सख्यात्मक मानों के माध्य का परिकलन किया जाय तो हमें एक ऐसी राशि प्राप्त होती है जिसका प्रयोग प्रसार के माप के लिए किया जा सकता है। इस माप को माध्य विचलन (*mean deviation*) कहते हैं।

$$\text{माध्य विचलन} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \quad (2 \text{ I I})$$

यहाँ $|x_i - \bar{x}|$ के अर्थ हैं $(x_i - \bar{x})$ और $(\bar{x} - x_i)$ में से वह राशि जिसका मान धनात्मक (*positive*) हो। अथवा यदि बारंबारता सारणी से परिकलन करना हो तो—

$$\text{माध्य विचलन} = \frac{\sum_{i=1}^k f_i |x_i - \bar{x}|}{\sum_{i=1}^k f_i} \quad (2 \text{ I 2})$$

उदाहरण

$$\begin{aligned} \text{सारणी सख्या 2.2 में } \bar{x} &= 26.15 \text{ वर्ष होने के कारण माध्य विचलन} \\ &= \frac{1}{20} [(3.15 \times 1) + (1.15 \times 8) + (0.15 \times 2) \\ &\quad + (0.85 \times 4) + (1.85 \times 5)] \text{ वर्ष} \\ &= \frac{1}{20} [3.15 + 9.20 + 0.30 + 3.40 + 9.25] \text{ वर्ष} \end{aligned}$$

$$\text{माध्य विचलन} = 1.265 \text{ वर्ष}$$

(घ) जब सब प्रेक्षणों का उनके परिमाणों के अनुसार विन्यास किया जाता है तो मध्य के प्रेक्षण को माध्यिका कहते हैं। इसी प्रकार वह प्रेक्षण जिससे 25 प्रतिशत प्रेक्षण छोटे और 75 प्रतिशत प्रेक्षण बड़े होते हैं—प्रथम-चतुर्थांश (*first quartile*)

कहलाता है। जिस प्रेक्षण से 75 प्रतिशत अवलोकन छोटे और 25 प्रतिशत प्रेक्षण बड़े होते हैं वह तृतीय चतुर्थक कहलाता है। द्वितीय चतुर्थक स्वयं माध्यिका होता है।

तृतीय चतुर्थक और प्रथम चतुर्थक के अंतर को अंतश्चतुर्थक-परास (*inter-quartile range*) कहते हैं। यह भी प्रसार का एक माप है।

परिमाणों के अनुसार विन्यास में जैसे 25-25 प्रतिशत प्रेक्षणों के अंतर पर चतुर्थक होते हैं उसी प्रकार दस दस प्रतिशत के अंतर पर दशमक (*decile*) तथा एक एक प्रतिशत के अंतर पर शततमक (*percentile*) होते हैं। दशमको तथा शततमको द्वारा प्रायः संपूर्ण वितरण का भास हो जाता है। परंतु जब तक बारंबारता चित्र न बनाया जाय तब तक इन सौ मापों से तत्व को पाना इतना ही कठिन हो जाता है जितना कि कुल प्रेक्षणों से। इसलिए केंद्रीय प्रवृत्ति तथा प्रसार के मापों के अतिरिक्त दो और माप कर्तुवता (*Kurtosis*) और वैपम्य होते हैं जिनसे हमें वितरण को समझने में सहायता मिलती है।

§ २८ घूर्ण (Moments)

इसके पूर्व कि हम इन दो मापों का वर्णन करें, आइए आपको एक समुदाय से परिचित कराया जाय जिसके दो सदस्यों से आप पहिले ही परिचय प्राप्त कर चुके हैं। इस समुदाय के सदस्यों को घूर्ण (*moment*) कहते हैं। यदि हम किसी वितरण के समस्त घूर्णों को जान लें तो उसके विषय में और अधिक जानने योग्य बहुत कम रह जाता है। वितरण के r के घूर्णों को μ_r से सूचित करते हैं और इसकी परिभाषा निम्नलिखित सूत्र द्वारा होती है।

$$\mu_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r \quad \dots\dots\dots(2.13)$$

जहाँ कुल प्रेक्षणों की संख्या n है, x_i i वाँ प्रेक्षण है और \bar{x} प्रेक्षणों का माध्य है। इस प्रकार के घूर्णों को जो माध्य के अन्तरो से सबधित है माध्यातरिक घूर्ण (*moment about the mean*) कहते हैं। इसी प्रकार किसी और मान a के अन्तरो से सबधित घूर्णों को a -आतरिक घूर्ण कहते हैं और इसे μ'_r से सूचित करते हैं।

$$\mu'_r = \frac{1}{n} \sum_{i=1}^n (x_i - a)^r \quad \dots\dots\dots(2.14)$$

माध्यमिक घूर्णों को a -आंतरिक घूर्णों के रूप में रखा जा सकता है।

$$\begin{aligned} \mu_r &= \sum_{i=1}^n (x_i - \bar{x})^r \\ &= \sum_{i=1}^n [(x_i - a) - (\bar{x} - a)]^r \\ &= \sum_{i=1}^n (x_i - a)^r - \binom{r}{1} (\bar{x} - a) \sum_{i=1}^n (x_i - a)^{r-1} \\ &\quad + \binom{r}{2} (\bar{x} - a)^2 \sum_{i=1}^n (x_i - a)^{r-2} + \dots + (-1)^r n (\bar{x} - a)^r \end{aligned}$$

$$\text{अथवा } \mu_r = \mu_{r(a)} - \binom{r}{1} (\bar{x} - a) \mu_{r-1(a)} + \binom{r}{2} (\bar{x} - a)^2 \mu'_{r-2(a)} + \dots + (-1)^r (\bar{x} - a)^r \quad (2.15)^*$$

यह आप समझ ही सके होंगे कि शून्यान्तरिक प्रथम घूर्ण

$$\mu'_1 = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

$$\text{तथा } \mu_2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

दस माध्यान्तरिक द्वितीय घूर्ण को प्रसरण (variance) कहते हैं।

आप इन दो घूर्णों से पहिले से ही परिचित हैं।

§ २.९ वैषम्य और ककुदता

दो मुख्य लक्षण जो वितरण के रूप की व्याख्या करते हैं (१) वैषम्य (skewness) या असममिति (asymmetry) तथा (२) ककुदता (kurtosis) या शिखरता (peakedness) हैं। इन दो लक्षणों के माप क्रमशः β_1 और β_2 हैं। इनकी परिभाषा निम्नलिखित सूत्रों से होती है।

* फुटनोट --- (r_1) , (r_2) इत्यादि की परिभाषा के लिए देखिए समीकरण (3.15)

$$\beta_1 = \frac{\mu'_3}{\mu'_2} \dots \dots \dots (216)$$

$$\beta_2 = \frac{\mu'_4}{\mu'^2_2} \dots \dots \dots (217)$$

उदाहरण — मारणी सख्या 2 2 2

$$\begin{aligned} \mu'_3(25) &= \frac{1}{20} [\{(-2)^3 \times 1\} + \{(1)^3 \times 2\} + \{(2)^3 \times 4\} + \{(3)^3 \times 5\}] (\text{वर्ष})^3 \\ &= \frac{1}{20} [-8 + 2 + 32 + 135] (\text{वर्ष})^3 \end{aligned}$$

$$= \frac{161}{20} (\text{वर्ष})^3$$

$$= 8 \ 05 (\text{वर्ष})^3$$

$$\mu'_4(25) = \frac{1}{20} [\{(-2)^4 \times 1\} + \{(1)^4 \times 2\} + \{(2)^4 \times 4\} + \{(3)^4 \times 5\}] (\text{वर्ष})^4$$

$$= \frac{1}{20} [16 + 2 + 64 + 405] (\text{वर्ष})^4$$

$$= \frac{487}{20} (\text{वर्ष})^4$$

$$= 24 \ 35 (\text{वर्ष})^4$$

यह हम पहिले ही कलम कर चुके है कि

$$\mu'_2(25) = 3 \ 35 (\text{वर्ष})^2$$

और $\bar{x} = 25$ वर्ष $= 1 \ 15$ वर्ष

$$\therefore \mu_2 = \mu'_2 - (\bar{x} - 25)^2$$

$$= [3 \ 35 - (1 \ 15)^2] (\text{वर्ष})^2$$

$$= 2 \ 0275 (\text{वर्ष})^2$$

$$\mu_3 = [\mu'_3 - 3\mu'_2(\bar{x} - 25) + 2(\bar{x} - 25)^2] (\text{वर्ष})^3$$

$$= [\{8 \ 05\} - \{3 \times 3 \ 35 \times 1 \ 15\} + \{2 \times (1 \ 15)^3\}] (\text{वर्ष})^3$$

$$= [8 \ 05000 - 11 \ 557500 + 2 \ 841730] (\text{वर्ष})^3$$

$$= -0 \ 665770 (\text{वर्ष})^3$$

$$\mu_4 = [\mu'_4 - 4\mu'_3(\bar{x} - 25) + 6\mu'_2(\bar{x} - 25)^2 - 3(\bar{x} - 25)^4] (\text{वर्ष})^4$$

$$= [\{24 \ 35\} - \{4 \times 8 \ 05 \times 1 \ 15\} + \{6 \times 3 \ 35 \times (1 \ 15)^2\}]$$

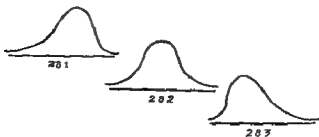
$$\begin{aligned}
 &= [3 \times (115)^3] (\text{वर्ग})^4 \\
 &= [24\ 35 - 37\ 03 + 26\ 58225 - 4\ 90198425] (\text{वर्ग})^4 \\
 &= 9\ 00026575 (\text{वर्ग})^4
 \end{aligned}$$

$$\begin{aligned}
 \therefore \beta_1 &= \frac{\mu_3^2}{\mu_2^3} = \frac{(0\ 66577)^2}{(2\ 0275)^3} \\
 &= 0\ 0531821
 \end{aligned}$$

$$\begin{aligned}
 \beta_2 &= \frac{\mu_4}{\mu_2^2} = \frac{9\ 00026575}{(2\ 0275)^2} \\
 &= 2\ 189442
 \end{aligned}$$

यह आसानी से देखा जा सकता है कि यदि वितरण सममित ((*symmetrical*) हो यानी किसी भी परिमाण a के लिए प्रेक्षणों के मान $(\bar{x}-a)$ तथा $(a-\bar{x})$ ग्रहण करने की बारंबारता बराबर हो—तो सभी विषम घूर्णों (odd moments) का मान शून्य होगा। इस कारण असममिति को मापने के लिए μ_3 उपयुक्त प्रतीत होता है। परन्तु इसको माप के मापक (unit) से स्वतन्त्र करने के लिए हम इसके वर्ग को μ_2^3 से विभाजित कर देते हैं। इस प्रकार असममिति का माप β_1 एक संख्या है जिसका कोई मापक नहीं है। जितना अधिक β_1 का मान होया वितरण उतना ही अधिक असममिति होगा। यह असममिति किस प्रकार की है यह जानने के लिए बजाए β_1 के इसके वर्ग मूल को लेना अधिक उत्तम है जिसका चिह्न μ_3 का चिह्न लिया जाय। इस वर्ग मूल को γ_1 से सूचित किया जाता है।

$$\begin{aligned}
 \gamma_1 &= \sqrt{\beta_1} \\
 &= \frac{\mu_3}{\mu_2^{3/2}}
 \end{aligned}$$



चित्र १०—असममित तथा सममित वितरण

ऊपर के उदाहरण में आपने यह देखा ही होगा कि μ_3 का मान उन प्रेक्षणों पर अधिक निर्भर करता है जो माध्य से अधिक अंतर पर हों। यदि इस प्रकार के प्रेक्षणों में माध्य से बड़े प्रेक्षणों की बारबारता अधिक हो तो वितरण का रूप उस प्रकार का होगा जैसा चित्र संख्या १० (281) में दिखाया गया है और इस दशा में μ_3 का और इसी कारण γ_2 का मान ऋणात्मक होता है। इसके विपरीत यदि माध्य से अधिक अंतर के प्रेक्षणों में माध्य से छोटे प्रेक्षणों का बाहुल्य हो तो वितरण का रूप चित्र १० (283) में दिये हुए बारबारता चित्र की तरह होगा। इस दशा में γ_2 का मान ऋणात्मक होगा। इस प्रकार γ_2 के मान से बाराबारता चित्र के रूप पर काफी प्रकाश पड़ता है।

$$\text{ककुदता का माप } \beta_2 = \frac{\mu_4}{\mu_2^2}$$

$$\begin{aligned} \text{परंतु } \mu_4 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 \\ &= \frac{1}{n} \sum_{i=1}^n \left[\{(x_i - \bar{x})^2 - \mu_2\} + \mu_2 \right]^2 \\ &= \frac{1}{n} \left[\sum_{i=1}^n \{(x_i - \bar{x})^2 - \mu_2\}^2 + 2\mu_2 \sum_{i=1}^n \{(x_i - \bar{x})^2 - \mu_2\} + n\mu_2^2 \right] \\ &= \mu_2^2 + \frac{1}{n} \sum_{i=1}^n \{(x_i - \bar{x})^2 - \mu_2\}^2 \\ \therefore \sum_{i=1}^n \{(x_i - \bar{x})^2 - \mu_2\} &= 0 \\ \therefore \beta_2 &= 1 + \frac{1}{n} \sum_{i=1}^n \left[\left(\frac{x_i - \bar{x}}{\sigma} \right)^2 - 1 \right]^2 \\ &= 1 + V \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 \end{aligned}$$

जहाँ $V \left(\frac{x_i - \bar{x}}{\sigma} \right)^2$ से हमारा तात्पर्य $\left(\frac{x_i - \bar{x}}{\sigma} \right)^2$ के प्रसरण (variance) से है

और $\sigma^2 = \mu_2$ । यह प्रसरण जितना कम या अधिक होगा उतना ही कम या अधिक β_2 का मान होगा । यह देखा गया है कि जिन वटनों के लिए β_2 अधिक होता है उनमें बारबारता चित्र माध्य के पास अधिक चपटा सा होता है और जिनमें इसका मान कम होता है उसमें यह माध्य के पास सिखर का सा रूप लिए होता है । प्रसामान्य वटन (*normal distribution*) में—जिसका वर्णन आगे के अध्यायों में किया जायगा—इसका मान 3 होता है । इसके बारबारता चित्र से तुलना करके यह अंदाजा लगाया जा सकता है कि एक विशिष्ट कबूदता वाले वटन का रूप माध्य के पास क्या होगा । β_2 की इस प्रकार की व्याख्या वास्तव में युक्तिपूर्ण नहीं है, फिर भी सांख्यिकी के साहित्य में इसका एक विशिष्ट स्थान है ।

प्रायिकता

§ ३१ वे स्थितियाँ जिनमें प्रायिकता का प्रयोग किया जाता है

पहिले अध्याय में कुछ ऐसी स्थितियों का वर्णन किया गया था जिनमें निश्चय पूर्वक किसी घटना की भविष्यवाणी करना सम्भव नहीं है। यह बतहा गया था कि ऐसी स्थितियों में सांख्यिकीय नियमों का उपयोग किया जाता है। ये अधिकतर प्रायिकता के रूप में होते हैं। इस अध्याय में हम प्रायिकता से परिचय प्राप्त करेंगे।

उन सब स्थितियों में जहाँ प्रायिकता का प्रयोग किया जाता है एक विशेषता पायी जाती है। आवश्यक है कि हम इस विशेषता को ध्यान में रखें, उदाहरणार्थ जुए के खेलों में, इश्योरेंस की समस्याओं में तथा पानी के बरसने में। हम देखते हैं कि ये सब घटनाएँ बार-बार घटने वाली हैं। पाँस का फेंकना एक ऐसी घटना है जो कब से कब कल्पना में तो अनगिनत बार पुहरायी जा सकती है, यदि हम इस समय इस सम्भावना की उपेक्षा करें कि पाँस घिस अथवा टूट जायगा। यदि हम इश्योरेंस की किसी एक लाक्षणिक समस्या को सुलझाने में लगे हैं तो हम कल्पना कर सकते हैं कि लाखों मनुष्य एक ही प्रकार का इश्योरेंस करवायेंगे और इन मनुष्यों से सम्बन्धित समान घटनाओं को इश्योरेंस कम्पनी के रजिस्ट्रारों में बोट कर लिया जायगा। पानी बरसने के समय में हम अनगिनत दिनों की कल्पना कर सकते हैं जो गुजर चुके हैं अथवा भविष्य में आनेवाले हैं। किन्तु हर एक दिन किसी विशेष स्थान पर कितनी वर्षा हुई होगी, यही वह घटना है जिसमें हमें रुचि है। सामूहिक घटनाओं का—जो प्रायिकता के प्रयोग के लिए उपयुक्त हैं—एक अच्छा उदाहरण है कुछ गुणों की वंशानुक्रमिता। किसी विशेष जाति के पौधों को ही लीजिए जो प्रारंभ में एक ही बीज से उत्पन्न हुए हो और उनके फूलों का रंग निरीक्षण करिए। यहाँ हम आसानी से समझ सकते हैं कि बार-बार घटित होने वाली घटनाएँ क्या हैं। विशेष रूप से एक पौधे का लगाना और उसके फूलों के रंगों का निरीक्षण करना केवल यही एक घटना है।

इसके पदचान् हम इस प्रकार की हजारों घटनाओं का केवल फूला के रंग के दृष्टिकोण से विश्लेषण करते हैं।

पाँसे फेंकने में प्रारम्भिक घटना पाँसे को एक बार फेंकना और जितने विट्टे ऊपर के पाखंड पर आयें उन्हें नोट कर लेना है। हैड और टेल के खेल में रुपये की प्रत्येक टॉम या उछाल एक घटना है और जो मुख ऊपर की ओर आये वही इस घटना का गुण (attribute) है। जीवन के बीमों में किसी एक व्यक्ति का जीवन एक घटना है और जिस युग का निरोक्षण किया जाता है वह है उस व्यक्ति का मृत्यु के समय की उम्र अथवा वह उम्र जिस पर बीमा कर्मी को उस मनुष्य अथवा उसके घर वाला को रुपया देना पड़ता है। जब हम एक मनुष्य की एक विशेष समय-अंतराल के अंदर मरने की प्रायिकता के बारे में बात करते हैं तो इसका एक विशेष अर्थ होता है। हमें किसी व्यक्ति विशेष नहीं बरन् व्यक्तियों के एक पूरे समुदाय के बारे में विचार करना होता है। उदाहरण के लिए यह समुदाय उन सब व्यक्तियों का हो सकता है जिनकी उम्र पचास वर्ष की हो और जिन्होंने जीवन का बीमा करा रखा हो। प्रायिकता की जो परिभाषा हम देंगे वह एक समूह में एक गुण के पाये जाने की बारबारता से ही संबंधित है। यदि आप यह कहते हैं कि बरकतउल्लाह के एक वर्ष के अन्दर ही मर जाने की प्रायिकता पचास प्रतिशत है तो इसका अर्थ केवल यह है कि बरकत-उल्लाह एक ऐसे समुदाय का सदस्य है जिसमें से पचास प्रतिशत व्यक्ति एक वर्ष के अन्दर ही मर जायेंगे। यह ध्यान में रखने की बात है कि यह वक्तव्य बरकतउल्लाह से कम और उस समुदाय से अधिक संबंधित है जिसका बरकतउल्लाह एक सदस्य है।

§ ३२ आपेक्षिक बारबारता का सीमान्त मान

मान लीजिए, एक सिपाही बन्दूक से निशाना लगाने का अभ्यास कर रहा है। उसने दो सी गज के अंतर पर एक तस्ला लगा रखा है जिसके बीच में एक ऊर्ध्व (vertical) रेखा खिंची हुई है। वह उस रेखा पर निशाना बाँधकर गोली चलाता है। कुछ गोलियाँ इस रेखा के बायीं ओर पड़ती हैं और कुछ दाहिनी ओर। इस क्रम में कोई नियम नहीं है। यह नहीं है कि बारी-बारी से गोलियाँ दाहिनी ओर बायीं ओर पड़ें या हर एक गोली के बाद जो बायीं भाग पर पड़ती है दो गोलियाँ दाहिनी ओर पड़ेंगी। वास्तव में इसमें किसी प्रकार का नियम दृष्टिगोचर नहीं होता। इस अभ्यास में प्रथम पन्द्रह गोलियाँ बिना किस जगह पड़ीं, यह चिन ११ में दिखाया गया है। क्या इस ज्ञान से हमें यह भविष्यवाणी करने में कुछ भी सहायता मिलती है

कि अगली गोली दाहिने भाग में पड़ेगी अथवा बायें भाग में ? प्रत्यक्ष है कि इस प्रकार की कोई भविष्य बाणी बरना समभव नहीं है। इस अनियमितता के होते हुए भी इस प्रयोग के फलों में कुछ नियम है। यदि सिपाही अच्छा निशानेबाज हो तो हम देखेंगे कि हजारों गोलियाँ चलाने के बाद करीब आधे निशान बायाँ ओर और आधे निशान दाहिनी ओर होंगे। यदि वह अच्छा निशानेबाज न भी हो और यदि हम हर गोली के पड़ने के बाद दाहिनी ओर पड़नेवाली गोलियों की बारबारता का और कुल गोलियों की संख्या का अनुपात निकालें तो हम देखेंगे कि जैसे जैसे कुल संख्या बढ़ती जाती है वैसे वैसे यह आपेक्षिक बारबारता (*relative frequency*) एक विशेष संख्या की ओर अभिसर होती जाती है। इस प्रकार विशेष संख्या की ओर अभिसर होने के क्या अर्थ हैं, यह भली भाँति समझना आवश्यक है।

	9°	•1	•10
		•8	
	2°		•4
3°			
		•5	
	11°		
12°		•6	
	14°		•75
		7°	
	15°		

चित्र ११—ऊर्ध्व रेखा पर निशाना बाँककर चलायी हुई गोलियों का वितरण

मान लीजिए कि आप इस आपेक्षिक बारबारता का परिकलन एक विशेष दशमलव स्थान तक करते हैं। यदि यह परिकलन पहिले दशमलव स्थान तक करना हो तो उदाहरण के लिए तीस में से दस गोलियाँ दाहिनी ओर पड़ने पर यह आपेक्षिक बार-

बारता $\frac{10}{30} = 0.3$ होगी। आप देखेंगे कि लगभग 500 गोलियाँ चलानेके बाद इस पहिले दशमलव स्थान तक परिकलित आपेक्षिक बारबारता का मान स्थिर हो जाता है और फिर चाहे कितनी ही अधिक गोलियाँ क्यों न चलायी जायें यह मान 0.5 ही बना रहता है। यदि आप दो दशमलव स्थानों तक इस आपेक्षिक बारबारता का परिकलन करें तो कदाचित् इस हजार निशानों के बाद यह 0.50 पर स्थिर हो जायगी। यदि तीन दशमलव स्थानों तक यह परिकलन किया जाय तो कई लाख प्रयोगों के पश्चात् यह स्थिर हो जायगी। किसी भी दशमलव स्थान तक परिकलन किया जाय प्रयत्नों की एक विशेष सख्या के परचान् यह स्थिरता आ ही जाती है। इन निरीक्षणों से हम इस निष्कर्ष पर पहुँचते हैं कि आपेक्षिक बारबारता एक विशेष सख्या की ओर प्रवृत्त होती है और जैसे जैसे प्रयत्नों की सख्या बड़ती जाती है आपेक्षिक बारबारता इस विशेष सख्या के अधिकाधिक पास आती जाती है।

हम लोग प्रायिकता के सिद्धान्तों में केवल उन बार-बार घटनेवाली घटनाओं के समुदायों का अध्ययन करेंगे जिनमें यह विदवास करने के काफी कारण हों कि आपेक्षिक बारबारता एक विशेष सख्या की ओर प्रवृत्त होती है। इस सख्या को आपेक्षिक बारबारता की सीमा (limit) कहते हैं। यह सीमा ही समुदाय में उस गुण के पाने जाने की प्रायिकता (probability) कहलानी है जिसकी आपेक्षिक बारबारता का परिकलन हम कर रहे थे।

५ ३ ३ एक अन्य परिभाषा

इस प्रायिकता शब्द की एक और परिभाषा है जो नीचे लिखे उदाहरणों द्वारा स्पष्ट हो जायगी।

(१) डिब्बा और गोलियाँ—एक डिब्बे में n_1 गोलियाँ हैं जिनमें n_2 सफेद हैं और बाकी अन्य दूसरे रंगों की। हम एक गोली को बिना देखे ही डिब्बे में से निकालते हैं, उसके रंग को नोट करते हैं और फिर उसे डिब्बे में वापस रख देते हैं। यह प्रयोग हम बार-बार करते हैं और अनगिनत बार कर सकते हैं। इन प्रयोगों में सफेद गोलियों की आपेक्षिक बारबारता जिस सीमा की ओर प्रवृत्त हो रही है उसे (ऊपर दी हुई परिभाषा के अनुसार) हम सफेद गोलियों के चुने जाने की प्रायिकता कहेंगे। परन्तु यदि रंग के अलावा गोलियाँ बनावट और वजन में समान हों और गोलियों को हर प्रयोग के बाद भली भाँति मिला दिया जावे तो यह स्वाभाविक जान पड़ेगा कि किसी भी गोली के चुने जाने की प्रायिकता उतनी ही है जितनी किसी अन्य गोली की। क्योंकि कुल n

गोलियाँ हैं जिनमें से n_1 गोलियाँ सफेद हैं, इसलिए सफेद गोली के चुने जाने की प्रायिकता $\frac{n_1}{n}$ है। अब प्रायिकता की परिभाषा यह भी मानी जा सकती है कि

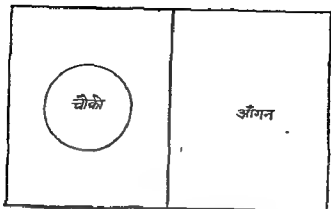
$$\text{प्रायिकता} = \frac{\text{विभिन्न एक-भी घटनाओं की संख्या}}{\text{समस्त विभिन्न घटनाओं की संख्या}} \quad (31)$$

यहाँ पर ऐसी घटनाओं पर विचार किया जा रहा है जिनकी प्रायिकताएँ सहज ज्ञान द्वारा (intuitively) समान मानी जा सकती हैं। यह आपने देखा होगा कि हम परिभाषा में प्रायिकता का कुछ ज्ञान पहिले से निहित है। इस कारण परिभाषा के रूप में यह उचित प्रतीत नहीं होगी। वास्तव में यदि समस्त प्राथमिक घटनाओं (elementary events) की प्रायिकता बराबर हों तो यह सूत्र केवल किसी समुच्चय घटना की प्रायिकता का कलन करने का एक नियम बताता है। ऊपर के प्रयोग में किसी एक गोली का निकालना एक प्राथमिक घटना है और सब प्राथमिक घटनाओं की प्रायिकताओं का बराबर मान लेना विचार-संगत मालूम होता है। किन्तु सफेद गोली का चुनाव एक समुच्चय घटना (joint event) है जो उन प्राथमिक घटनाओं के समुच्चय से बनी है जिनमें विभिन्न सफेद गोलियों का चुनाव होता है।

यह भी स्पष्ट हो है कि प्रेक्षण द्वारा प्रायिकता का पता लगाना असंभव है, क्योंकि इसके लिए अज्ञेय प्रयोग करने पड़ेंगे। अगले अध्याय में हम देखेंगे कि प्रायिकता किस सिद्धान्त के आधार पर निश्चित की जाती है। प्रेक्षण द्वारा हमें यह मालूम हो सकता है कि यह निश्चित प्रायिकता संभव है या नहीं। ऐसी परिस्थितियों में जहाँ प्राथमिक घटनाओं की प्रायिकता बराबर जान पड़ती है हमारा प्रयोग करना अनावश्यक प्रतीत होता है।

(२) बर्षा—मान लीजिए, आप एक छोटे-से आँगन में खड़े हैं। उसमें एक चौकी पड़ी है। थोड़ी देर में हलकी हलकी फुहारें पड़ने लगती हैं। इतनी हलकी कि आप हर बूंद को—जो आँगन में गिरती है—गिन सकते हैं और यह भी देख सकते हैं कि वह चौकी पर गिरी या नहीं। खास बूंदों के गिरने के बाद आप उस प्रायिकता का किसी हद तक अनुमान लगा सकते हैं जो कि किसी बूंद के चौकी पर गिरने की है। यह अनुमान आप चौकी पर गिरी हुई बूंदों की आपेक्षिक बारंबारता के आधार पर लगायेंगे। यदि वर्षा जोरों से पड़ रही है तो बूंदों का गिनना असंभव है।

यदि आप आँगन को उसकी मुजाबों से समानांतर रेखाओं द्वारा छोटे छोटे किन्तु बराबर क्षेत्रफलवाले वर्गों (squares) में विभाजित कर दें तो ऊपर के उदा-



चित्र १२—चौकी पर वर्गा-बिन्दुओं की प्रायिकता

हरण की भाँति यहाँ भी यह विचार समत मालूम होता है कि प्रत्येक वर्ग में बूँद के पडने की प्रायिकता बराबर है।

∴ बूँद के चौकी पर पडने की प्रायिकता

$$\frac{\text{उन वर्गों की संख्या जो चौकी में हैं}}{\text{कुल वर्गों की संख्या जो पूरे आंगन में हैं}}$$

परंतु कुछ वर्ग ऐसे भी हैं जो अक्षत चौकी पर और अक्षत उसके बाहर हैं। यदि इन वर्गों की संख्या उन वर्गों की अपेक्षा बहुत कम है जो चौकी में हैं तो प्रायिकता के बलन में ऊपर के सूत्र के प्रयोग से कोई विरोध अंतर नहीं पडेगा। मान लीजिए पूरे आंगन में पाँच करोड़ वर्ग हैं जिनमें से एक करोड़ चौकी पर पूर्णतया और एक सत्स्र अक्षत पडते हैं। इस दशा में हम कह सकते हैं कि यदि बूँद के चौकी पर पडने की प्रायिकता वास्तव में p है तो

$$p < \frac{10,000,000 + 1,000}{50,000,000} = +\frac{1}{5} + \frac{1}{50,000,000}$$

$$\text{और } p > \frac{10,000,000}{50,000,000} = \frac{1}{5}$$

(‘ k ’ > ‘ x ’ के अर्थ होते हैं कि ‘ x ’ से ‘ k ’ बड़ा है। इसी प्रकार ‘ k ’ < ‘ x ’ के अर्थ होते हैं कि ‘ x ’ से ‘ k ’ छोटा है।)

इस प्रकार हमने बूँदों के चौकी पर पडने की प्रायिकता की दो सीमाएँ निश्चित कर ली और हम यह कह सकते हैं कि प्रायिकता इन दोनों सीमाओं के बीच की कोई

सख्या है। यदि हम अधिकाधिक छोटे वर्ग लेते चले जायें तो ये सीमाएँ भी पास आती जायेंगी। सीमान्त में दोनों बराबर हो जायेंगी। सीमान्त में चौकी पर स्थित वर्गों की सख्या का कुल वर्गों की सख्या से अनुपात चौकी और आंगन के क्षेत्रफल के अनुपात के बराबर होता है। इस प्रकार—

$$\text{बूंद के चौकी पर गिरने की प्रायिकता} = \frac{\text{चौकी का क्षेत्रफल}}{\text{आंगन का क्षेत्रफल}}$$

किसी भी मौसम विज्ञान विभाग (meteorological station) में वर्षा को नापने के लिए जो वृष्टि-मापक (rain-gauge) लगाया जाता है उसमें इस ऊपर लिखे सिद्धान्त का प्रयोग किया जाता है। उस वृष्टि मापक में जितना पानी पड़ता है उसे बाहर में पड़े हुए पानी का प्रतिनिधि मानने में यही तर्क है।

§ ३४ प्रतिवधी प्रायिकता

किसी घटना अथवा गुण की प्रायिकता के लिए यह भी आवश्यक है कि हम यह जानें कि वह किस प्रयोग से सम्बन्धित है। उदाहरणार्थ, ऊपर हम चौकी पर बूंद गिरने की प्रायिकता का परिकलन कर रहे थे। इसमें प्रयोग था उन बूंदों का निरीक्षण जो आंगन में गिर रही हैं। यदि आंगन के बीच में एक रेखा खींची हुई हो और हम केवल उन बूंदों का निरीक्षण करें जो रेखा के उस ओर वाले भाग में गिर रही हैं जिसमें चौकी है तो बूंद के चौकी पर गिरने की प्रायिकता बदल जायेगी। वास्तव में हमें यह कहना चाहिए कि उन बूंदों के लिए जो पूरे आंगन में गिर रही हैं चौकी पर गिरने की प्रायिकता चौकी और आंगन के क्षेत्रफलों के अनुपात के बराबर है।

इसी प्रकार यदि हम रुपया उछालते हैं और देखते हैं कि वह चित गिरता है या गढ़ तो एक अच्छे सिक्के के लिए चित गिरने की प्रायिकता $\frac{1}{2}$ है। इस प्रयोग में समस्त उत्क्षेपणों (tosses) के परिणामों का निरीक्षण किया जाता है। प्रयोग को बदल कर यह प्रतिवधि लगाया जा सकता है कि हम केवल उन उत्क्षेपणों पर विचार करेंगे जिनके पूर्वगामी उत्क्षेपण का परिणाम पट हो। मान लीजिए कि प्रथम सोलह उत्क्षेपणों के परिणाम निम्नलिखित हैं—

1	2	3	4	5	6	7	8
चि	चि	प	चि	प	प	प	चि
		=	=	=	=	=	

9 10 11 12 13 14 15 16

प चि प प चि प चि प
= = = =

इसमें हम केवल चौथे, छठे, सातवें, आठवें, दसवें, बारहवें, तेरहवें, तथा पंद्रहवें उत्क्षेपणों पर आपेक्षिक बारबारता के परिवर्तन के लिए विचार करेंगे, क्योंकि ये ही उत्क्षेपण पट पडने के पश्चात् के हैं। इस प्रकार की आपेक्षिक बारबारता को प्रतिबन्धी आपेक्षिक बारबारता (*conditional relative frequency*) कहते हैं। इस विशेष उदाहरण में हम यह कहेंगे कि यह दिये हुए होने पर कि पिछले उत्क्षेपण का परिणाम पट या चित पडने की प्रतिबन्धी आपेक्षिक बारबारता $\frac{2}{5}$ है।

इस प्रकार की प्रतिबन्धी आपेक्षिक बारबारता की सीमा को प्रतिबन्धी प्रायिकता कहते हैं।

§ ३५ स्वतंत्र घटनाएँ

मान लीजिए कि A और B दो घटनाएँ हैं। यदि A की प्रायिकता बिना किसी प्रतिबन्ध के उतनी ही हो जितनी इस प्रतिबन्ध के साथ कि B उससे पहिले घटित हो चुकी है, तो हम कहते हैं कि घटना A घटना B से स्वतंत्र है।

आगे से हम किसी घटना A की प्रतिबन्धहीन प्रायिकता को $P(A)$ द्वारा सूचित करेंगे। इसी प्रकार A की प्रतिबन्धी प्रायिकता को—यह दिया होने पर कि B घटित हो चुकी है $-P(A/B)$ द्वारा सूचित किया जायगा और इसे प्रायिकता A दत्त B' पडा जाता है।

इस संकेत (notation) के अनुसार A घटना B से स्वतंत्र कहलायेगी यदि

$$P(A/B) = P(A)$$

§ ३६ घटनाओं का संगम और प्रतिच्छेद (Intersection)

किसी एक ही प्रयोग के परिणाम स्वरूप कई भिन्न-भिन्न घटनाएँ हो सकती हैं। इन्हें हम प्राथमिक घटनाएँ (elementary events) कह सकते हैं। कुछ और घटनाएँ ऐसी होती हैं जो इनमें से कुछ विशेष प्राथमिक घटनाओं का कुलम (set) होती हैं। उदाहरण के लिए एक पाँसे की फेंकने से 1, 2, 3, 4, 5 अथवा 6 बिंदु ऊपर आ सकते हैं। इस प्रकार यह छ तो प्राथमिक घटनाएँ हैं। किन्तु केवल 1, 3 या 5 में से किसी भी एक संख्या का ऊपर आना इस प्रकार की घटनाओं का एक कुलम है। प्रायिकता की भाषा में इस प्रकार की घटनाओं को प्राथमिक घटनाओं का संगम

(union) कहते हैं। यदि A और B दो घटनाएँ हो तो हम इनके सगम का सांकेतिक निरूपण $A \cup B$ के द्वारा करते हैं और इसे 'A सगम B' पढ़ते हैं। इसका शाब्दिक अर्थ है A या B में से कम से कम एक घटना का घटित होना।

एक और प्रकार की घटना A और B से मवधित हो सकती है। यह है A और B दोनों का एक साथ घटित होना। मान लीजिए कि एक रुपये को दो बार उछाला जाता है। घटना A पहिले उरक्षेपण में रुपये का चित पडना है और घटना B है दूसरे उरक्षेपण में चित पडना। यदि दोनों उरक्षेपणों में रुपया चित आये तो A भी घटित होगी और B भी। इस प्रकार दो घटनाओं A और B के एक साथ घटित होने को हम A और B का प्रतिच्छेद कहते हैं। इसको $A \cap B$ द्वारा सूचित करते हैं, और इसे 'A प्रतिच्छेद B' पढ़ते हैं।

§ ३'७ परस्पर अपवर्जी घटनाएँ (Mutually Exclusive Events)

कुछ घटनाएँ ऐसी होती हैं जो साथ-साथ हो ही नहीं सकती। जैसे पाँसा फेकने पर १ और २ दोनों साथ साथ ऊपर नहीं आ सकते। इस प्रकार की घटनाओं को परस्पर अपवर्जी घटनाएँ कहते हैं। यदि A और B दो परस्पर अपवर्जी घटनाएँ हैं तो $A \cap B$ एक ऐसी घटना है जो हो ही नहीं सकती। ऐसी असंभव घटनाओं को हम \emptyset द्वारा सूचित कर सकते हैं।

इस प्रकार यदि हम लिखें कि—

$$A \cap B = \emptyset$$

तो इसका अर्थ यह होगा कि A और B परस्पर अपवर्जी घटनाएँ हैं।

§ ३'८ घटनाओं का वियोग

मान लीजिए प्रयोग पासे को फेकने का है और A तथा B निम्नलिखित घटनाएँ हैं।

A: 1, 2 या 3 बिंदुओं में से किसी एक का ऊपर आना

B: 2, 4 या 6 बिंदुओं में से किसी एक का ऊपर आना

इस दशा में A और B का सगम निम्नलिखित है।

$A \cup B$: 1, 2, 3, 4 या 6 बिंदुओं का ऊपर आना। इसी प्रकार A और B का गुणनफल निम्नलिखित है

$A \cap B$: 2 बिंदुओं का ऊपर आना।

यदि १ अथवा ३ बिंदु ऊपर आगें तो A घटित होगी परंतु B नहीं। इस प्रकार की घटना को हम $A-B$ से सूचित करते हैं और इसे "A वियोग B" कहते हैं। इसी प्रकार यदि B घटित हो और A नहीं तो इसको $B-A$ से सूचित करते हैं। ऊपर की घटनाओं के लिए

$A-B$: १ अथवा ३ बिंदुओं का ऊपर आना

$B-A$: ४ अथवा ६ बिंदुओं का ऊपर आना

§ ३.९ घटनाओं का गभित होना

मान लीजिए ऊपर के प्रयोग में एक घटना C है।

C: २ अथवा ३ बिंदुओं में से किसी एक का ऊपर आना।

यह स्पष्ट है कि यदि C घटित होगी तो A भी घटित होगी। इसको हम संकेत द्वारा निम्नलिखित तरीके से सूचित करते हैं

$$C \subset A$$

शब्दों द्वारा हम यह कह सकते हैं कि 'घटना C घटना A में गभित है'।

आप यह आसानी से देख सकते हैं कि—

$$(A \cap B) \subset A$$

$$(A \cap B) \subset B \quad \dots\dots\dots(3.2)$$

$$(A-B) \subset A$$

$$(B-A) \subset B$$

यदि कोई घटना C घटना A में गभित नहीं हो तो इस गुण को संकेत द्वारा हम निम्नलिखित रीति से सूचित कर सकते हैं :

$$C \not\subset A$$

§ ३.१० आपेक्षिक बारंबारता के कुछ गुण

एक बात शायद आपके ध्यान में आयी होगी। वह यह कि जहाँ भी हम घटनाओं के अनन्त अनुक्रम (infinite sequence) अथवा बारंबारता के सीमान्त मानों का वर्णन करते हैं वहाँ हम केवल विचारों की दुनिया में विचरण कर रहे हैं। वास्तव में किसी भी मनुष्य को घटनाओं के अनन्त अनुक्रम का निरीक्षण नहीं करना होता और बारंबारताओं के सीमांत मानों का कोई भौतिक अस्तित्व नहीं है। आप कदाचित् सोचते होंगे कि इस प्रकार की धारणा का व्यावहारिक जीवन में क्या उपयोग हो सकता

है। परंतु प्रयोजित गणित (applied mathematics) इस प्रकार की धारणाओं से भरा हुआ है। उदाहरण के लिए गति-विज्ञान (dynamics) में किसी एक बिंदु पर वेग (velocity) अथवा किसी एक बिंदु पर त्वरण (acceleration) इस प्रकार की धारणाएँ हैं जिनका भौतिक अस्तित्व नहीं है और न उनका प्रेक्षण किया जा सकता है। वास्तव में ये किसी अल्प समय-अंतराल में वर्तमान वेग अथवा त्वरण के सीमान्त मान ही हैं। परंतु हम जानते हैं कि इन्हीं धारणाओं को आधार स्वरूप लेकर जो गतिविज्ञान निर्मित हुआ है उसका उपयोग इंजीनियर लोग करते हैं। यद्यपि इनका अपना अस्तित्व नहीं है, परंतु ये कुछ ऐसे गुणों का आदर्शिकरण (idealisation) हैं जो वास्तविक हैं। इसी प्रकार यद्यपि प्रायिकता भी एक सीमान्त मान है परंतु वह उस आपेक्षिक बारंबारता से गन्धित है जिसके भौतिक अस्तित्व को हम पहिचानते हैं।

भाइए, अब हम आपेक्षिक बारंबारताओं के कुछ गुणों से परिचय प्राप्त करें, क्योंकि जिस प्रायिकता का हमें अध्ययन करना है उसमें भी ये गुण अवश्य ही विद्यमान रहेंगे।

(१) यदि n प्रयोगों में किसी घटना की बारंबारता ν हो तो $\frac{\nu}{n}$ इस घटना की आपेक्षिक बारंबारता हुई। यह स्पष्ट है कि ν न तो शून्य से कम कोई ऋणात्मक संख्या हो सकती है और न यह n से अधिक ही हो सकता है। इस कारण आपेक्षिक बारंबारता न तो ऋणात्मक संख्या हो सकती है और न १ से अधिक कोई घनात्मक संख्या। आपेक्षिक बारंबारताओं के इस गुण को सूत्र में हम लिख सकते हैं

$$0 \leq \frac{\nu}{n} \leq 1 \quad \dots\dots (3.3)$$

(२) यदि कोई घटना असंभव हो तो बारंबारता ν शून्य होगी। इस कारण असंभव घटनाओं की आपेक्षिक बारंबारता भी शून्य होगी।

(३) यदि किसी घटना का प्रयोग के साथ होना अनिवार्य हो तो $\nu = n$ होगा तथा इस दशा में घटना की आपेक्षिक बारंबारता १ होगी।

आगे से हम किसी विशेष घटना A की बारंबारता को $\nu(A)$ द्वारा सूचित करेंगे।

(४) यदि A और B परस्पर अपवर्जी घटनाएँ हों जिनकी आपेक्षिक बारंबारताएँ क्रमशः $\nu(A)$ और $\nu(B)$ हों तो इन दोनों घटनाओं के संगम $A \cup B$ की आपेक्षिक बारंबारता $\nu(A) + \nu(B)$ होगी। इस गुण को हम निम्नलिखित सूत्र द्वारा सूचित कर सकते हैं

यदि $A \cap B = 0$ हो तो,

$$\nu(A \cup B) = \nu(A) + \nu(B) \quad \dots\dots (3.4)$$

(५) यदि $\nu(A|B)$ B के घट चुकने पर A की प्रतिवधी-आपेक्षिक बारबारता को सूचित करता है तो

$$\nu(A|B) = \frac{\nu(A \cap B)}{\nu(B)} \quad (35)$$

क्योंकि मान लीजिए कि B की बारबारता ν_2 , AUB की बारबारता ν , और कुल बारबारता n है।

$$\text{तो } \nu(B) = \frac{\nu_2}{n}$$

$$\nu(A \cap B) = \frac{\nu'}{n}$$

$$\begin{aligned} \text{तथा } \nu(A|B) &= \frac{\nu'}{\nu} \\ &= \frac{\nu' / n}{\nu_2 / n} \\ &= \frac{\nu(A \cap B)}{\nu(B)} \end{aligned}$$

§ ३११ प्रायिकता के गुण

क्योंकि प्रायिकता आपेक्षिक बारबारता का सीमान्त मान है, इसलिए उसके गणो और आपेक्षिक बारबारता के ऊपर लिखे गुणों में समानता होनी आवश्यक है। यही नहीं प्रायिकता की एक परिभाषा जो आजकल सबसे अधिक मान्य है निम्न-लिखित है—

प्रायिकता यादृच्छिक प्रयोगों (random experiments) के परिणामों से संबंधित एक माप है जिसके निम्नलिखित गुण हैं—

(१) यदि A एक असंभव घटना है तो $P(A) = 0$

(२) यदि A एक अनिवार्य घटना है तो $P(A) = 1$

(१, २) P एक माप है जिसका निम्नतम मान शून्य और महत्तम मान 1 है अथवा

$$0 \leq P(A) \leq 1 \quad (36)$$

(३) यदि A और B दो परस्पर अपवर्ती घटनाएँ हों तो

$$P(A \cup B) = P(A) + P(B) \quad (37)$$

(3') इसी प्रकार यदि $A_1, A_2, A_3, \dots, A_n$ कुल n परस्पर अपवर्जी घटनाएँ हों तो

$$P(\bigcup_{i=1}^n A_i) = P(A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i) \quad (3.8)$$

(3') यदि A_1, A_2, \dots इत्यादि अनगिनत अपवर्जी घटनाएँ हों तो इनके संगम को $\bigcup_{i=1}^{\infty} A_i$ से सूचित किया जा सकता है और

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \quad (3.9)$$

(4) यदि $P(B) > 0$ न हो तो B के दिय होने पर A की प्रतिवधी प्रायिकता का नीचे लिखे सूत्र द्वारा परिवर्तन किया जा सकता है

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (3.10)$$

(4) गुणन का नियम यदि A_1, A_2, \dots, A_n कुल n घटनाएँ हों तो

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3 \cap \dots \cap A_n) &= P(A_1 | A_2 \cap A_3 \cap \dots \cap A_n) P(A_2 \cap A_3 \cap \dots \cap A_n) \\ P(A_1 \cap A_2 \cap \dots \cap A_n) &= P(A_2 | A_3 \cap A_4 \cap \dots \cap A_n) P(A_3 \cap A_4 \cap \dots \cap A_n) \\ P(A_1 \cap A_2 \cap \dots \cap A_n) &= P(A_3 | A_4 \cap A_5 \cap \dots \cap A_n) P(A_4 \cap A_5 \cap \dots \cap A_n) \end{aligned}$$

$$\begin{aligned} P(A_1 \cap A_2 \cap \dots \cap A_n) &= P(A_n | A_n) P(A_n) \\ P(A_1 \cap A_2 \cap \dots \cap A_n) &= P(A_n) P(A_{n-1} | A_n) P(A_{n-2} | A_{n-1} \cap A_n) \\ &\quad P(A_{n-3} | A_{n-2} \cap A_{n-1} \cap A_n) \end{aligned} \quad (3.11)$$

(5) यदि A और B दो स्वतंत्र घटनाएँ हों तो परिभाषा के अनुसार

$$P(A \cap B) = P(A)P(B)$$

परन्तु चौथे गुण के अनुसार

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$\text{इसलिए } P(A \cap B) = P(A)P(B)$$

(3.12)

(5') इसी प्रकार यदि A_1, A_2, \dots, A_n परस्पर स्वतंत्र घटनाएँ हों तो

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2) \dots P(A_n) \dots \dots \dots (3.13)$$

आइए, अब हम ऊपर दी हुई धारणाओं से अधिक परिचित होने के लिए प्रायिकता की कुछ प्रहेलिकाओं को हल करें।

प्रहेलिकाएँ

(१) घुडदौड़ में दौब लगाने की आम प्रथा है। एक प्रकार की घुडदौड़ में सात घोड़े दौड़ते हैं और यदि आप उनके क्रम की ठीक-ठीक भविष्यवाणी कर दें तो आपको एक सहस्र रुपये का लाभ होता है। यदि आप घोड़ों के बारे में कुछ नहीं जानते और केवल अनुमान के आधार पर भविष्यवाणी करते हैं तो क्या प्रायिकता है कि आपको यह सहस्र रुपये की प्राप्ति हो जायेगी ?

यदि हम सात भिन्न भिन्न वस्तुओं के कुल क्रमचयों ((permutations)) की संख्या को $7!$ से सूचित करें तो प्रायिकता का कलन निम्नलिखित विधि से हो सकता है

(31) के अनुसार

$$\begin{aligned} \text{प्रायिकता} &= \frac{\text{विभिन्न अनुकूल घटनाओं की संख्या}}{\text{समस्त विभिन्न घटनाओं की संख्या}} \\ &= \frac{\text{उन क्रमचयों की संख्या जिनके चुनाव पर आपको लाभ होगा}}{\text{कुल क्रमचयों की संख्या}} \\ &= \frac{1}{7!} \end{aligned}$$

यदि A, B, C और D चार विभिन्न वस्तुएँ हैं तो उनको निम्नलिखित क्रमों में सजाया जा सकता है।

(1) ABCD	(7) BACD	(13) CABD	(19) DABC
(2) ABDC	(8) BADC	(14) CADB	(20) DACB
(3) ACBD	(9) BCDA	(15) CBAD	(21) DBAC
(4) ACDB	(10) BCAD	(16) CBDA	(22) DBCA
(5) ADBC	(11) BDAC	(17) CDAB	(23) DCAB
(6) ADCB	(12) BDCA	(18) CDBA	(24) DCBA

जिस प्रकार ऊपर के उदाहरण में सात वस्तुओं के कुल क्रमचयों की संख्या को $7!$ से सूचित किया था, उसी प्रकार हम चार वस्तुओं के कुल क्रमचयों की संख्या को $4!$ से सूचित करते हैं। यहाँ हम देख ही चुके हैं कि

$$4! = 24 \\ = 4 \times 3 \times 2 \times 1$$

इसी प्रकार यदि n विभिन्न वस्तुओं के क्रमचयों की संख्या को $n!$ से सूचित किया जाय तो यह सिद्ध किया जा सकता है कि

$$n! = n \times (n-1) \times (n-2) \times \dots \times 3 \times 2 \times 1 \quad (3.14)$$

इस प्रकार ऊपर के उदाहरण में

$$\text{प्रायिकता} = \frac{1}{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1} \\ = \frac{1}{5,040}$$

इसके अर्थ यह हुए कि यदि इस प्रकार की घुड़दौड़ों में आप बारबार क्रम के मक्खन में भविष्यवाणी करे तो औसतन 5040 भविष्यवाणियों में से एक ठीक होगी। यह बात आपने नोट की होगी कि इस भविष्यवाणी के प्रयोग में प्रत्येक क्रमचय एक संभव प्रायिकता घटना है। ये सब प्रायिकता घटनाएँ परस्पर अपवर्जी हैं और हमने यह भाग लिया है कि इन सब क्रमचयों की चुने जाने की प्रायिकता समान है। यह कल्पना इस स्थान पर उचित ही प्रतीत होती है।

(२) एक कारखाने में बिजली के बल्ब बनते हैं जिनमें औसतन सौ में से पाँच खराब निकल जाते हैं। यदि दिन भर के उत्पादन में जो लाखों बल्ब हैं उनमें से हम यादृच्छिक विधि से 4 बल्ब चुन लेते हैं तो इन चुने हुए बल्बों में से 3 के खराब होने की क्या प्रायिकता है ?

हम किसी ऐसे क्रमचय के चुनने की प्रायिकता का विचार करें जिसमें 3 बल्ब खराब हों। यदि हम अच्छे बल्बों को A से और बुरे बल्बों को B से सूचित करें तो एक क्रमचय निम्नलिखित हो सकता है।

ऐसे क्रमचय को चुनने की प्रायिकता

B B B A

$= P$ [पहिले बल्ब का बुरा होना \cap दूसरे बल्ब का बुरा होना \cap तीसरे बल्ब का बुरा होना \cap चौथे बल्ब का अच्छा होना]

$$= (\text{पहिले बल्ब के बुरे होने की प्रायिकता}) \times \\ (\text{दूसरे बल्ब के बुरे होने की प्रायिकता}) \times \\ (\text{तीसरे बल्ब के बुरे होने की प्रायिकता}) \times$$

(चौथे बल्ब के अच्छे होने की प्रायिकता)

$$= \frac{5}{100} \times \frac{5}{100} \times \frac{5}{100} \times \frac{95}{100}$$

$$= \frac{19}{160,000}$$

यह परिकलन इस कल्पना के आधार पर किया गया है कि यह सयुक्त घटना जिन चार घटनाओं का गुणनफल है वे स्वतंत्र हैं। यहाँ समीकरण (3 13) का उपयोग किया गया है।

इस प्रकार हम देखेंगे कि तीन बुरे और एक अच्छे बल्ब के जितने भी क्रमचय हैं उनकी प्रायिकता $\frac{19}{160,000}$ है। ऐसे कुल क्रमचय चार हैं।

(1) BBBA (2) BBAB (3) BABB (4) ABBB

यह चारों परस्पर अपवर्जी घटनाएँ हैं। इसलिए इसकी प्रायिकता कि इनमें से कोई भी एक घटित हो जाय समीकरण (3 8) के अनुसार

$$P[(BBBA) \cup (BBAB) \cup (BABB) \cup (ABBB)]$$

$$= P(BBBA) + P(BBAB) + P(BABB) + P(ABBB)$$

$$= \frac{19}{160,000} + \frac{19}{160,000} + \frac{19}{160,000} + \frac{19}{160,000}$$

$$= \frac{76}{160,000} = \frac{19}{40,000}$$

यदि कुल N वस्तुएँ हों जिनमें से r एक प्रकार की और $(N-r)$ दूसरे प्रकार की हो तो समस्त क्रमचयों की संख्या को—जो एक दूसरे से भिन्न हो— $\binom{N}{r}$ से सूचित किया जाता है। इस संकेत का प्रयोग हम पिछले अध्याय में कर चुके हैं। ऊपर के उदाहरण में $N=4$ और $r=1$

$$\therefore \text{कुल विभिन्न क्रमचयों की संख्या} = \binom{4}{1}$$

$$= 4$$

यह सिद्ध किया जा सकता है कि

$$\binom{N}{r} = \frac{N!}{r!(N-r)!} \quad (3 15)$$

उदाहरण के लिए यदि चार बल्बों में से दो बुरे और दो अच्छे हों तो कुल प्रयोजनों की संख्या

$$\begin{aligned} \binom{4}{2} &= \frac{4!}{2!2!} \\ &= \frac{4 \times 3 \times 2 \times 1}{2 \times 1 \times 2 \times 1} \\ &= 6 \end{aligned}$$

ये गिन कर भी देखे जा सकते हैं

- | | |
|----------|----------|
| (1) AABB | (4) BBAA |
| (2) ABAB | (5) BABA |
| (3) ABBA | (6) BAAB |

ऐसे क्रमचयों को जिनमें एक ही प्रकार की विभिन्न वस्तुओं में भेद नहीं किया जाता, मध्य (Combination) कहते हैं।

(३) ऊपर के ही उदाहरण में इस घटना की क्या प्रायिकता है कि चुने हुए चार बल्बों में से कम से कम एक बल्ब अच्छा हो ?

यहाँ दो परस्पर अपवर्जी घटनाएँ हैं

- (क) कम से कम एक बल्ब अच्छा हो।
 (ख) चारों बल्ब खराब हों।

इसके अतिरिक्त और कोई घटना संभव नहीं है।

अर्थात् इन दोनों में से एक घटना का होना निश्चित है।

प्रायिकता के दूसरे गुण के कारण

$$\therefore P[(\text{कमसे कम एक बल्ब अच्छा हो}) \cup (\text{चारों बल्ब खराब हों})] = 1$$

परन्तु इस समीकरण में बायीं ओर का भाग

$$= P[\text{कम से कम एक बल्ब अच्छा हो}]$$

$$+ P[\text{चारों बल्ब खराब हो}]$$

$$\therefore P[\text{कमसे कम एक बल्ब अच्छा हो}]$$

$$= 1 - P[\text{चारों बल्ब खराब हो}]$$

परन्तु $P[\text{चारों बल्ब खराब हो}] = P(B B B B)$

$$= \frac{5}{100} \times \frac{5}{100} \times \frac{5}{100} \times \frac{5}{100}$$

$$= \frac{1}{160,000}$$

$$\therefore P[\text{कम से कम एक बल्ब अच्छा हो}] = \frac{159,999}{160,000}$$

(४) तास के पत्ता में से दो पत्ते C_1 और C_2 खींचे गये। हम A से इस घटना को सूचित करेंगे कि C_1 पान का पत्ता है और B से इस घटना को कि C_2 पान का पत्ता है।

$$\text{स्पष्टतया समीकरण (3 I) के अनुसार } P(A) = \frac{13}{52}$$

यदि हमें पता हो कि A घटित हो चुकी है तो C_2 बाकी के 51 पत्तों में से यादृच्छिक विधि द्वारा खींचा गया एक पत्ता है। इन पत्तों में केवल 12 पत्ते पान के हैं। इसलिए

$$\text{समीकरण (3 I) के अनुसार } P(B|A) = \frac{12}{51}$$

इस बात की प्रायिकता कि दोनों पत्तों पान के हैं प्रायिकता के गुणन के नियम समीकरण (3 II) के अनुसार $P(A \cap B) = P(A)P(B|A)$

$$= \frac{13}{52} \times \frac{12}{51}$$

$$= \frac{1}{17}$$

६.३.१२ बेज का प्रमेय (Bayes' Theorem)

गुणन नियम के अनुसार

$$\begin{aligned} P(A \cap B) &= P(A)P(B|A) \\ &= P(B)P(A|B) \end{aligned}$$

$$\therefore P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (3 I6)$$

मान लीजिए कि $A_1, A_2, A_3, \dots, A_n$ कुल n परस्पर अपवर्जी घटनाएँ हैं जिनका B के साथ हो सकना मगभव है।

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup \dots \cup (A_n \cap B)$$

$$\therefore P(B) = P \left[\bigcup_{p=1}^n (A_p \cap B) \right]$$

$$= \sum_{\nu=1}^n P(A_{\nu} \cap B)$$

यदि $P(A_{\nu}) = \pi_{\nu}$, तब, $P(B|A_{\nu}) = P_{\nu}$
 $\nu = 1, 2, 3, \dots, n$

तो

$$\begin{aligned} P(A_{\nu}/B) &= \frac{P(A_{\nu})P(B/A_{\nu})}{P\left[\bigcup_{\nu=1}^n (A_{\nu} \cap B)\right]} \\ &= \frac{P(A_{\nu})P(B/A_{\nu})}{\sum_{\nu=1}^n P(A_{\nu} \cap B)} = \frac{P(A_{\nu})P(B/A_{\nu})}{\sum_{\nu=1}^n P(A_{\nu})P(B/A_{\nu})} \\ &= \frac{P_{\nu} \pi_{\nu}}{\sum_{\nu=1}^n P_{\nu} \pi_{\nu}} \end{aligned} \quad (3.17)$$

यह सूत्र बेज का प्रमेय कहलाता है।

इस प्रमेय का प्रयोग बहुधा निम्नलिखित अवस्था में होता है। किसी एक यादृच्छिक प्रयोग में हम घटना B के होने अथवा न होने का निरीक्षण करते हैं। हमें यह पता है कि A_1, A_2, \dots, A_n कुल n परस्पर अपवर्जी कारण हैं जिनके फलस्वरूप घटना B हो सकती है। मान लीजिए कि प्रयोग के पहिले ही हमें यह मालूम हो जाता है कि कारण A_{ν} के प्रभावकारी होने की प्रायिकता क्या है। इसको A_{ν} की पूर्वत गृहीत प्रायिकता (a-priori probability) कहते हैं। मान लीजिए कि यह पूर्वत गृहीत प्रायिकता $P(A_{\nu}) = \pi_{\nu}$ है। परंतु A_{ν} के प्रभावकारी होने पर भी यह आवश्यक नहीं है कि घटना B घटे ही। मान लीजिए कि B की प्रतिवधी प्रायिकता $P(B/A_{\nu}) = P_{\nu}$ है, जब प्रतिवध यह हो कि A_{ν} काय कर रहा है।

बेज के प्रमेय के आधार पर हम A_{ν} की प्रायिकता $P(A_{\nu}/B)$ का परिकलन कर सकते हैं। यानी B के प्रेक्षण के पश्चात् हम A_{ν} के प्रभावकारी होने की प्रायिकता मालूम कर सकते हैं। इसे A_{ν} की परत लब्ध प्रायिकता (a-posteriori probability) कहते हैं।

सांख्यिकी में इस प्रमेय के उपयोग में सबसे बड़ी बाधा यह है कि अधिकतर पूर्वत गृहीत प्रायिकता अज्ञात होती है। नीचे हम एक छोटा सा उदाहरण देते हैं जहाँ इस प्रमेय का युक्तियुक्त प्रयोग हो सकता है।

उदाहरण—पाँच बर्तन हैं जिनमें से हर एक में चार-चार गोलियाँ हैं। इन बर्तनों को पृथक् पृथक् पहिचानने के लिए हम इनका नामकरण संस्कार करके इन्हें A_1, A_2, A_3, A_4 तथा A_5 कहेंगे। इनमें दो रंग की गोलियाँ हैं—नीली और लाल। किस बर्तन में कितनी गोलियाँ लाल और कितनी नीली हैं यह नीचे दिया हुआ है।

- A_1 — चारो नीली गोलियाँ
 A_2 — तीन गोलियाँ नीली और एक लाल।
 A_3 — दो गोलियाँ नीली और दो लाल।
 A_4 — एक गोली नीली और तीन लाल।
 A_5 — चारो लाल गोलियाँ।

प्रयोग के पहिले भाग में एक बर्तन यादृच्छिक विधि से चुना जाता है। फिर चुने हुए बर्तन में से दो गोलियाँ यादृच्छिक विधि से चुनी जाती हैं। हर एक गोली को चुनने के बाद उसको वापस बर्तन में रख दिया जाता है। यदि दोनो चुनी हुई गोलियाँ लाल हों तो तीसरे चुनाव में भी पाँचो बर्तनों में से लाल गोली के चुने जाने की क्या प्रायिकता होगी ?

यदि हम दोनो गोलियो के लाल होने की घटना को B से सूचित करें तो

$$P(B) = \frac{\left(\frac{0}{4}\right)^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{2}{4}\right)^2 + \left(\frac{3}{4}\right)^2 + \left(\frac{4}{4}\right)^2}{5}$$

$$= \frac{30}{16 \times 5}$$

$$= \frac{3}{8}$$

$B \cap C$ द्वारा हम उस घटना को सूचित करते हैं जिसमें तीनो चुनी हुई गोलियो का रंग लाल हों।

$$P(B \cap C) = \frac{\left(\frac{0}{4}\right)^3 + \left(\frac{1}{4}\right)^3 + \left(\frac{2}{4}\right)^3 + \left(\frac{3}{4}\right)^3 + \left(\frac{4}{4}\right)^3}{5}$$

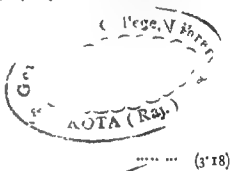
$$= \frac{100}{64 \times 5}$$

$$= \frac{5}{16}$$

$$\begin{aligned} \therefore P(C/B) &= \frac{P(B \cap C)}{P(B)} \\ &= \frac{5/16}{3/8} \\ &= 5/6 \end{aligned}$$

ऊपर के उदाहरण में यदि कुल $n+1$ बर्तन हो जिनमें से प्रत्येक में गोलियों की संख्या n और लाल गोलियों की संख्या क्रमशः $0, 1, 2, 3, \dots, n$ हो और यदि प्रथम n चुनावों में लाल गोलियाँ चुनी गयी हो तो $(n+1)$ वें चुनाव पर भी लाल गोली के चुने जाने की प्रायिकता

$$\begin{aligned} P &= \frac{\sum_{r=1}^n \left(\frac{r}{n}\right)^{n+1}}{\sum_{r=1}^n \left(\frac{r}{n}\right)^n} \\ &= \frac{n+1}{n+2} \end{aligned}$$



जहाँ $\frac{n+1}{n+2}$ के संकेत के अर्थ हैं लगभग बराबर होना।

इस सूत्र के प्रयोग के समय हमें यह बात ध्यान में रखनी चाहिए कि हमें यह ज्ञात है कि हर एक बर्तन के चुने जाने की प्रायिकता बराबर है। कुछ लोग इस सूत्र का प्रयोग खस अवस्था में भी करते हैं जब उन्हें इन प्रायिकताओं के बारे में कोई ज्ञान नहीं होता। ऐसी अज्ञान की अवस्था में वे विभिन्न सचयों की प्रायिकता को समान मान लेते हैं। परन्तु यह उपयोग उचित नहीं है।

लाप्लास ने इसका प्रयोग सूर्य के उदय होने की प्रायिकता के परिकलन के लिए किया था। यदि प्राचीन रिकार्डों के आधार पर हम यह जानते हैं कि सूर्य पिछले पाँच सहस्र वर्षों में रोज उदय होता रहा है तो

$$n = 1,826,213 \text{ दिन}$$

$$\therefore \text{सूर्य के कल उदय होने की प्रायिकता} = \frac{1,826,214}{1,826,215}$$

अब यह तय करना बाप ही के ऊपर छोडा जाना है कि इस प्रकार प्रायिकता का परिवर्तन किम हद तक उचित है। मूत्र (3 18) को जिन अभिवारणाओं के आधार पर निकाला गया था क्या वे इस उदाहरण के लिए सत्य हैं ? कुल 1, 826, 214 दिनों में से जिन दिनों में मूत्रोदय हुआ हो उनकी मर्यादा के लिए मान 0, 1, 2, 1, 826, 214 धारण करने की क्या कोई पूर्वन गृहीत प्रायिकताएँ हैं ? यदि नहीं तो इन्डि-नुमार इन प्रायिकताओं को समान समझ लेना वहाँ तक ठीक है ?

प्रायिकता वंटन और यादृच्छिक चर

(Probability Distribution and Random Variable)

५.४.१ यादृच्छिक चर

यादृच्छिक प्रयोग क्या होते हैं, यह आप जानते ही हैं। अधिकतर इन प्रयोगों के फलों को सख्या के रूप में रखा जा सकता है। जहाँ भी प्रयोग किसी चर के गिनने अथवा नापने से संबंधित है यह फल स्पष्टतया सख्या के रूप में रखे जा सकते हैं। कई और अवस्थाओं में भी हम सख्याओं से फलों को सूचित कर सकते हैं। उदाहरण के लिए एक नवजात शिशु के लिए हम एक संकेत बना सकते हैं जिसमें लड़के को 1 और लड़की को 0 से सूचित किया जाता हो। इसी प्रकार के नियम और अधिक जटिल परिस्थितियों में भी अपनाये जा सकते हैं।

इस अध्याय में और उसके पश्चात् भी हम अधिकतर उन्ही प्रयोगों के संबंध में चर्चा करेंगे जिनमें फल को सख्या का रूप दिया जा सकता हो। वह चर जो प्रयोग के फल को सूचित करता है यादृच्छिक चर (random variable) कहलाता है। यदि इस चरको X द्वारा सूचित किया जाय तो प्रयोग के भिन्न-भिन्न फलों के अनुसार X भिन्न-भिन्न मान धारण करता है। क्योंकि एक यादृच्छिक प्रयोग में विभिन्न फलों की निश्चित प्रायिकता होती है, इस यादृच्छिक चर X की विभिन्न मानों को धारण करने की प्रायिकता भी निश्चित हो जाती है।

$P(X=a)$ से हम उस घटना की प्रायिकता को सूचित करेंगे जब X का मान a हो। इसी प्रकार $P(a < X \leq b)$ द्वारा हम उस घटना की प्रायिकता को सूचित करेंगे जब कि X का मान a से अधिक और b से कम अथवा उसके बराबर हो। यदि हमें हर एक मान-युग्म a और b के लिए $P(a < X \leq b)$ ज्ञात हो तो हम कहते हैं कि हमें X का प्रायिकता वंटन (probability distribution) मालूम है।

उदाहरण के लिए पासे को फेंकने के यादृच्छिक प्रयोग को ही लीजिए। इसमें हम पाँसे के ऊपर के मुख पर बिंदुओं की सख्या को X से सूचित करेंगे। यह X एक

यादृच्छिक चर है जिसका मान 1, 2, 3, 4 5 और 6 हो सकता है। इन सब मानों की प्रायिकता बराबर है।

$$P(X=1)=P(X=2)=P(X=3)=P(X=4)=P(X=5)=P(X=6)=\frac{1}{6}$$

अब कोई भी दो सख्याएँ a और b को लेकर हम $P[a < X \leq b]$

का परिवर्तन सरलता से कर सकते हैं।

उदाहरणार्थ मान लीजिए $a=2$, $b=4$ 5।

$$\begin{aligned} P[a < X \leq b] &= P[2 < X \leq 4 \text{ 5}] \\ &= P[(X=3) \cup (X=4)] \\ &= P(X=3) + P(X=4) \\ &= \frac{1}{6} + \frac{1}{6} \\ &= \frac{1}{3} \end{aligned}$$

§ ४२ असतत वटन (Discrete distribution)

ऐसे वटन को जिसमें यादृच्छिक चर मानों की केवल एक परिमित (finite) सख्या धारण कर सकता है असतत वटन कहते हैं।

इस प्रकार का चर एक असतत चर कहलाता है। ऊपर के उदाहरण में यादृच्छिक चर X का वटन असतत है।

§ ४२१ यादृच्छिक चर के फलन का वटन

यदि X एक यादृच्छिक चर हो तो X का ऐसा फलन $g(X)$ भी जो X के किसी एक मान के लिए एक ही निश्चित मान धारण करता हो, एक यादृच्छिक चर है। ऊपर के उदाहरण के लिए X^2 एक यादृच्छिक चर है जिसका प्रायिकता वटन निम्नलिखित होगा

$$\begin{aligned} P[X^2=1] &= P[X=1] = \frac{1}{6} \\ P[X^2=1] &= P[X^2=4] = P[X^2=9] = P[X^2=16] = P[X^2=25] \\ &= P[X^2=36] = \frac{1}{6} \end{aligned}$$

क्योंकि X^2 एक चर है जिसके साथ एक प्रायिकता वटन संबन्धित है, इस कारण यह भी एक यादृच्छिक चर है। $Z = (X^2 - 3X)$ भी एक यादृच्छिक चर है जिसका प्रायिकतर वितरण निम्नलिखित विधि से गालूम किया जा सकता है।

यदि $[X=1]$ तो $\xi=1^2-3 \times 1=-2$

यदि $[X=2]$ तो $\xi=2^2-3 \times 2=-2$

यदि $[X=3]$ तो $\xi=3^2-3 \times 3=0$

यदि $[X=4]$ तो $\xi=4^2-3 \times 4=4$

यदि $[X=5]$ तो $\xi=5^2-3 \times 5=10$

यदि $[X=6]$ तो $\xi=6^2-3 \times 6=18$

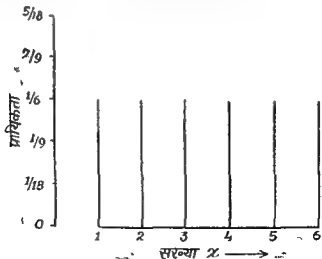
$\therefore P[\xi=-2]=P[(X=1) \cup (X=2)]=\frac{2}{6}$

और $P[\xi=0]=P[\xi=4]=P[\xi=10]=P[\xi=18]=\frac{1}{6}$

इस प्रकार X के किसी भी फलन का प्रायिकता-वंटन मालूम किया जा सकता है। यदि $g^{-1}(a, b)$ द्वारा हम X के उन सब मानों के कुलक (set) को सूचित करें जिनके लिए $a < g(X) \leq b$

तो $P[a < g(X) \leq b] = P[X \in g^{-1}(a, b)]$.. (4.1)

जहाँ $X \in g^{-1}(a, b)$ का अर्थ है X का $g^{-1}(a, b)$ में से कोई एक मान धारण करना। यदि हमें X का प्रायिकता वंटन ज्ञात है तो हम ऊपर के समीकरण में दाहिनी ओर के भाग का परिकलन कर सकते हैं। ऊपर के उदाहरण में



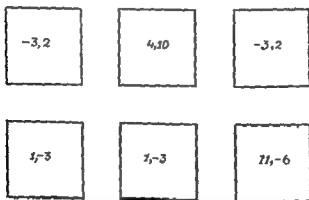
चित्र १३—पाँसा फेंकने पर ऊपर की बिन्दुओं की संख्या का प्रायिकता-वंटन

$$\begin{aligned}
 P[0 < X^2 \leq 5] &= P[0 < X \leq +\sqrt{5}] + P[-\sqrt{5} \leq X < 0] \\
 &= P[(X=1) \cup (X=2)] \\
 &= P(X=1) + P(X=2) \\
 &= \frac{1}{3}
 \end{aligned}$$

जिस प्रकार बारवारता बटन को चित्र द्वारा समझा जा सकता है उन्ही प्रकार प्रायिकता-बटन का भी चित्रण हो सकता है।

§ ४.२२ द्वि-विमितीय यादृच्छिक चर (Two-dimensional random variable)

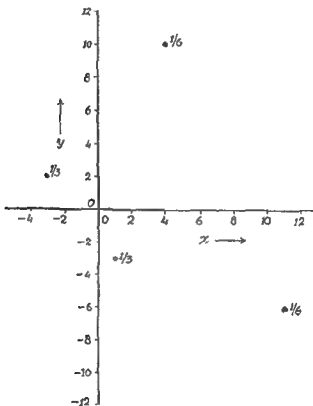
मान लीजिए कि एक पासा ऐसा बनाया गया है जिसके हर एक मुख पर दो सख्याएँ लिखी हुई हैं। प्रयोग है पासे को फेंककर ऊपर के मुख की सख्याओं को नोट करना।



चित्र १४—एक पासे के छः मुख

यह सख्याओं का युग्म एक यादृच्छिक चर है क्योंकि इसके भिन्न-भिन्न मानों के साथ प्रायिकता संबंधित है। इस प्रकार के चर को—जिसमें दो सख्याएँ किन्हीं विशेष क्रम में दी हुई हों—द्वि-विमितीय चर कहते हैं। जिस प्रकार अब तक हम यादृच्छिक चर को X से सूचित करते आये हैं उसी प्रकार एकद्वि-विमितीय चर को (X, Y) से सूचित किया जा सकता है। (X, Y) के प्रायिकता बटन की हम प्रायिकता द्रव्य-मान (Probability mass) की तरह कल्पना कर सकते हैं जो एक द्वि-विमितीय घरातल पर वितरित है। इसलिए इस प्रकार के बटन को चित्र द्वारा सूचित किया जा सकता

है। ऊपर के उदाहरण में जो (X, Y) का वंटन है उसे निम्न में नीचे दी हुई विधि से रखा जा सकता है।



चित्र १५—चित्र १४ में दिये हुए पासे को फेंकने से प्राप्त द्विविधतीय चर का वंटन

इस यादृच्छिक चर-युग्म के लिए

$$P [(X, Y) = (-3, 2)] = \frac{1}{3}$$

$$P [(X, Y) = (1, -3)] = \frac{1}{3}$$

$$P [(X, Y) = (4, 10)] = \frac{1}{6}$$

$$P [(X, Y) = (11, -6)] = \frac{1}{6}$$

§ ४.२.३ द्वि-विमितीय चर के फलन का वटन

हम देख चुके हैं कि यदि हमें X का प्रायिकता वटन ज्ञात हो तो हम उसके किसी भी फलन $g(X)$ का प्रायिकता वटन मालूम कर सकते हैं। इसी प्रकार यदि हमें (X, Y) का वटन ज्ञात हो तो इनके एक-मितीय तथा द्वि-मितीय फलनों के प्रायिकता वटन भी प्राप्त किये जा सकते हैं।

उदाहरण—यदि (X, Y) का वटन ऊपर लिखित है तो $P[(X+Y) \leq 10]$ क्या होगा ?

यदि $(X, Y) = (-3, 2)$ तो $(X+Y) = -1$

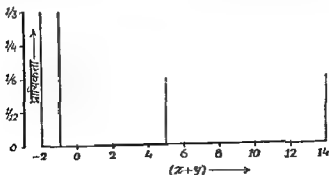
यदि $(X, Y) = (1, -3)$ तो $(X+Y) = -2$

यदि $(X, Y) = (4, 10)$ तो $(X+Y) = 14$

यदि $(X, Y) = (11, -6)$ तो $(X+Y) = 5$

$$\begin{aligned} \therefore P[(X+Y) \leq 10] &= P[\{(X+Y) = -1\} \cup \{(X+Y) = -2\} \\ &\quad \cup \{(X+Y) = 5\}] \\ &= P[(X+Y) = -1] + P[(X+Y) = -2] + P \\ &\quad [(X+Y) = 5] \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \\ &= \frac{3}{6} \end{aligned}$$

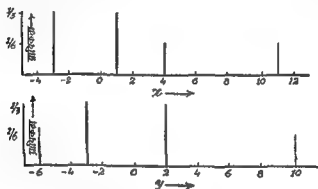
इसी प्रकार किन्हीं भी दो मानों a और b के बीच में $(X+Y)$ के पाये जाने की प्रायिकता का परिकलन भी किया जा सकता है। $(X+Y)$ एक विमितीय चर है जिसके प्रायिकता-वटन को निम्नलिखित रीति से चित्रित किया जा सकता है।



चित्र १६—चित्र १४ में दिये हुए पाँसे को फेंकने से प्राप्त ऊपर के मुख की संख्याओं के योग $(X+Y)$ का प्रायिकता-वटन

§ ४.२.४ एक-पार्श्वीय वंटन (*Marginal Distribution*)

(X, Y) का वंटन ज्ञात होने पर हम X और Y के वंटनों को अलग-अलग भी मालूम कर सकते हैं। इन वंटनों को एक-पार्श्वीय वंटन कहते हैं। ऊपर के चित्र, संख्या 15 में (X, Y) का वंटन दिखाया गया है। उसमें प्रायिकता द्रव्य-मान बिंदुओं का क्रमशः X और Y निर्देशाक्षों पर प्रक्षेप (projection) करने पर ये एक-पार्श्वीय वंटन प्राप्त हो सकते हैं।



चित्र १७—चित्र १५ में दिये हुए प्रायिकता-वंटन का निर्देशाक्षों पर विक्षेप— X और Y का एक पार्श्वीय वंटन।

यदि (X, Y) का वंटन ज्ञात हो तो हम X और Y के वंटन मालूम कर सकते हैं, परंतु यदि X और Y के वंटन मालूम हों तो (X, Y) का वंटन मालूम कर लेना संभव नहीं है। इसका कारण यह है कि (X, Y) के अनगिनत वंटन ऐसे मालूम किये जा सकते हैं जिनके एक-पार्श्वीय वंटन समान हों। उदाहरण के लिए (X, Y) के निम्नलिखित वंटनों का विचार कीजिए

- (1) $P[(X, Y)=(1, 1)] = \frac{1}{4}$
 $P[(X, Y)=(2, 1)] = \frac{1}{4}$
 $P[(X, Y)=(1, 2)] = \frac{1}{4}$
 $P[(X, Y)=(2, 2)] = \frac{1}{4}$
- (2) $P[(X, Y)=(1, 1)] = \frac{1}{8}$
 $P[(X, Y)=(2, 1)] = \frac{1}{8}$
 $P[(X, Y)=(1, 2)] = \frac{1}{8}$
 $P[(X, Y)=(2, 2)] = \frac{1}{8}$

इन दोनों द्वि-विभित्तीय वटनों के एक-पार्श्वीय वटन समान ही हैं जो निम्न-रूपित हैं—

$$X \text{ के लिए } P(X=1)=\frac{1}{2}, \quad P(X=2)=\frac{1}{2}$$

$$Y \text{ के लिए } P(Y=1)=\frac{1}{2}, \quad P(Y=2)=\frac{1}{2}$$

इससे यह सिद्ध हो गया कि X और Y दोनों के वटन ज्ञात होने पर भी समुक्त वटन (joint distribution) मालूम करना हमेशा सम्भव नहीं है। इसी प्रकार X और Y के एक-पार्श्वीय वटन मालूम होने से $(X+Y)$ का वटन मालूम कर लेना हमेशा सम्भव नहीं होता।

६.४.३ सतत वटन (Continuous distribution)

हम यह पहले ही कह चुके हैं कि किसी यादृच्छिक चर के प्रायिकता-वटन के ज्ञात होने का अर्थ है प्रत्येक मान युग्म a और b के बीच में इस चर के पाये जाने की प्रायिकता का ज्ञात होना। मान लीजिए कि हमें किसी यादृच्छिक चर X का वटन मालूम है। यदि x , δ और δ' कोई तीन संख्याएँ हैं तो हमें $P[a-\delta < X \leq a+\delta']$ अर्थात् X के अंतराल $[x-\delta, x+\delta']$ में पाये जाने की प्रायिकता ज्ञात होनी चाहिए।

इस अंतराल की लंबाई $(\delta+\delta')$ है और इस अंतराल में प्रायिकता $P[x-\delta < X \leq x+\delta']$ वितरित है। इसलिए औसतन अंतराल की एक इकाई लंबाई में प्रायिकता $\frac{P[x-\delta < X \leq x+\delta']}{\delta+\delta'}$ होगी। जिन दृष्टिकोण से प्रायिकता को

द्रव्य-मान के रूप में कल्पना की जा सकती, उसी दृष्टिकोण से ऊपर दी हुई यह औसत प्रायिकता प्रति इकाई अंतराल में प्रायिकता-घनत्व (probability density) समझा जा सकता है। δ और δ' के विभिन्न मानों के लिए हमें विभिन्न अंतराल प्राप्त होंगे और इनमें से प्रत्येक अंतराल के लिए प्रायिकता-घनत्व मालूम किया जा सकता है।

यदि δ और δ' के मानों को क्रमशः छोटे करते चले जायें, जिससे कि वे दोनों शून्य की ओर प्रवृत्त होने जायें, तो यह सम्भव है कि तत्संबन्धी अंतरालों में प्रायिकता-घनत्व किसी विशेष संख्या को ओर प्रवृत्त होता जाय। यदि ऐसा हो तो इस विशेष संख्या को हम यादृच्छिक चर X का बिंदु x पर प्रायिकता-घनत्व (probability density of the random variable X at point x) कहते हैं। इसी प्रकार दूसरे बिंदुओं पर केंद्रित अंतरालों में प्रायिकता घनत्व की सीमाएँ भी प्राप्त की जा सकती हैं।

आपका ध्यान कदाचिन् अपने पूर्व-परिचित चरो की ओर जायगा और आप यह जानना चाहेंगे कि इनके लिए विभिन्न बिंदुओं पर प्रायिकता घनत्व कितना है। वास्तव में अभी तक हमने जिन चरो से परिचय प्राप्त किया है वे गिनती में केवल थोड़े से ही मानों को धारण कर सकते हैं। अर्थात् दूसरे मानों के धारण करने की प्रायिकता इन चरो के लिए शून्य होती है।

मान लीजिए, हम एक ऐसा चर लेते हैं जिसके लिए

$$P(X=1)=P(X=2)=P(X=3)=P(X=4)=\frac{1}{4}$$

मान लीजिए x को 1 3 δ को 0 2 तथा δ' को 0 3 ले। तो इस अंतराल में प्रायिकता-घनत्व = $\frac{P[(1 3 - 0 2) < X \leq (1 3 + 0 3)]}{0 2 + 0 3}$

$$= \frac{P[1 1 < X \leq 1 6]}{0 5}$$
 होया।

परन्तु $P[1 1 < X < 1 6] = 0$ क्योंकि 1 1 और 1 6 के बीच का कोई मान X ग्रहण नहीं कर सकता, इसलिए यह घनत्व शून्य हुआ। अब यदि x को 1 3 ही रखा जाय तथा δ और δ' को क्रमश घटाते जायें तो आप देखेंगे कि इस प्रकार से प्राप्त प्रत्येक अंतराल में प्रायिकता-घनत्व शून्य होगा। इसलिए बिंदु $x=1 3$ पर X का प्रायिकता-घनत्व शून्य है। इसी प्रकार 1, 2, 3 और 4 को छोड़कर किसी भी बिंदु पर प्रायिकता-घनत्व शून्य होगा, यह सिद्ध किया जा सकता है।

आइए, अब हम यह देखें कि इन चार बिंदुओं पर प्रायिकता-घनत्व क्या है। मान लीजिए कि—

$$x=1 0, \delta=0 5, \delta'=0 5 \text{। } (x-\delta, x+\delta) \text{ में}$$

$$\text{प्रायिकता-घनत्व} = \frac{P[0 5 < X \leq 1 5]}{1 0}$$

$$= \frac{P(X=1)}{1 0}$$

$$= 1/4$$

यदि $x = 1 0, \delta = 0 2, \delta' = 0 2$ तो $(x-\delta, x+\delta')$ में

$$\text{प्रायिकता-घनत्व} = \frac{P[0 8 < X \leq 1 2]}{0 4}$$

$$\frac{P[X=1]}{0.4}$$

$$= \frac{1}{1.6}$$

यदि $x = 1.0$, $\delta = 0.01$, $\delta' = 0.01$ तो

$$\text{यह प्रायिकता-घनत्व} = \frac{P[X=1]}{0.02} = \frac{1}{0.08}$$

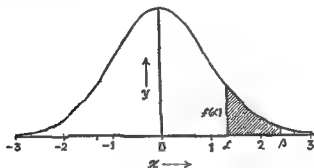
इस प्रकार हम देखते हैं कि ज्यों-ज्यों δ और δ' घटते जाते हैं त्यों-त्यों इस अनुपात में अंश (numerator) तो वही रहता है, परंतु हर (denominator) घटता चला जाता है। इस प्रकार δ और δ' को काफी छोटे मान लेकर इस अनुपात को हम किसी भी दिये हुए मान से अधिक बड़ा कर सकते हैं। इस प्रकार इस बिंदु पर प्रायिकता घनत्व अनंत है। इसी प्रकार बिंदु $x=2$, $x=3$ और $x=4$ पर भी प्रायिकता घनत्व अनंत सिद्ध किया जा सकता है। यह तो हमने एक उदाहरण लिया था, परंतु इसी प्रकार किसी भी असतत चर के लिए यह सिद्ध किया जा सकता है कि वह जिन मानों को किसी भी घनात्मक प्रायिकता से धारण कर सकता है उस पर उसका प्रायिकता-घनत्व अनंत और अन्य सब बिंदुओं पर उसका प्रायिकता-घनत्व शून्य होता है। इस प्रकार इन यादृच्छिक चरों के लिए विभिन्न बिंदुओं पर प्रायिकता-घनत्व जानने से हमें केवल यह मालूम हो सकता है कि किन बिंदुओं पर प्रायिकता शून्य नहीं है।

परंतु हम दूसरे अध्याय में सतत चरों से परिचय प्राप्त कर ही चुके हैं। यदि किसी यादृच्छिक प्रयोग द्वारा हमें इस प्रकार का चर प्राप्त हो तो यह एक सतत यादृच्छिक चर होगा। इस प्रकार के चर अपने परास में स्थित किसी भी दो मानों के बीच के सभी मानों को धारण कर सकते हैं। इस प्रकार के चर के लिए यदि हम इसके परास में कोई अंतराल लें तो स्पष्ट है कि इस पूरे अंतराल में चर के होने की प्रायिकता उस अंतराल के किसी भी छोटे भाग में होने की प्रायिकता से अधिक होगी। इस प्रकार किसी बिंदु पर केंद्रित अंतराल में प्रायिकता का परिफलन करते समय न केवल अंतराल की लंबाई शून्य की ओर प्रवृत्त होती है वरन् इस अनुपात का अंश (numerator) अर्थात् अंतराल में स्थित प्रायिकता भी शून्य की ओर प्रवृत्त होती है। इस प्रकार यह संभव है कि प्रायिकता-घनत्व शून्य और अनंत के बीच का कोई परिमित मान हो। इस प्रकार का बंटन जिसमें प्रत्येक बिंदु पर प्रायिकता-घनत्व अनंत से भिन्न कोई परिमित संख्या होगी है एक सतत बंटन कहलाता है।

यदि यादृच्छिक चर X का घटन सतत हो तो बिन्दु x पर इसके प्रायिकता घनत्व को $f(x)$ से सूचित करते हैं।

$$f(x) = \lim_{\delta \rightarrow 0} \frac{P[x - \delta < X \leq x + \delta]}{\delta + \delta'} \quad (4.2)$$

सतत घटन को हम बारबारता फलन $y=f(x)$ के ग्राफ या लेखा चित्र से चित्रित कर सकते हैं।



चित्र १८—एक सतत घटन का आवृत्ति फलन—

$$y=f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

इस वक्र और x -निर्देशाक्ष के बीच का क्षेत्रफल 1 होता है। यदि घनत्व-फलन $f(x)$ है तो यादृच्छिक चर X के अंतराल $[a, b]$ में पाये जाने की प्रायिकता

को $\int_a^b f(x) dx$ से सूचित किया जाता है। ऊपर के दिये हुए चित्र में X के

किसी मान x के लिए वक्र पर y का मान $f(x)$ है। यदि दो बिन्दुओं $(a,0)$ और $(b,0)$ से दो ऊर्ध्व रेखाएँ खींची जायें तो x -निर्देशाक्ष, बारबारता-वक्र और इन दो रेखाओं के बीच का क्षेत्रफल—जिसको चित्र में टेढ़ी रेखाओं से ढाँका हुआ है—

$\int_a^b f(x) dx$ ही होगा। इस प्रकार हमें इसचित्र द्वारा घटन का बहुत कुछ आभास

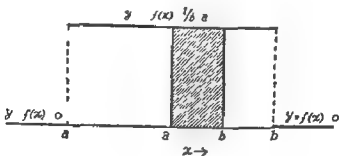
हो जाता है। इसका स्वरूप वही है जो समष्टि के लिए बारबारता-चित्र का होता है।

नाचे सतत वटनो के कुछ उदाहरण दिये हुए हैं ।

§ ४३१ आयताकार वटन (Rectangular distribution)

$$\begin{aligned}
 f(x) &= 0 && \text{यदि } x < a \\
 f(x) &= \frac{1}{b-a} && \text{यदि } a \leq x \leq b \\
 f(x) &= 0 && \text{यदि } x > b
 \end{aligned}$$

इस वितरण को आयताकार वटन (rectangular distribution) कहते हैं । इसका कारण यह है कि किन्हीं भी दो मानों के बीच में X के पाये जाने की प्रायिकता को एक आयत द्वारा चित्रित किया जा सकता है ।



चित्र १९—आयताकार वटन में $P[a < X \leq b]$

§ ४३२ प्रसामान्य वटन (Normal distribution)

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad -\infty < X < +\infty$$

जहाँ π एक वृत्त की परिधि (circumference) और व्यास (diameter) का अनुपात है तथा e एक संख्या है जिसका मान निम्नलिखित अनन्त श्रेणी (infinite series) से प्राप्त होता है ।

$$\begin{aligned}
 e &= 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \dots + \frac{1}{r!} + \dots \\
 &= 1 + \sum_{r=1}^{\infty} \frac{1}{r!}
 \end{aligned} \tag{4.3}$$

इस घटन का प्रायिकता घनत्व पहिले ही चिन सख्या १८ में चिनित किया जा चुका है।

यह स्पष्ट है कि किसी सतत घटन में चर के किसी भी मान a के लिए $P[X = a] = 0$ । यह इस कारण कि यह प्रायिकता ऊपर दिये हुए नियम के अनुमार दो ऊच्च रेखाओं के बीच का क्षेत्रफल होना चाहिए, परंतु जब इन दो रेखाओं के बीच का अंतर शून्य हो गया तो स्पष्ट है कि यह क्षेत्रफल भी शून्य होगा।

$$\text{अतः शब्दों में } \int_a^a f(x) dx = 0 \quad (4.4)$$

§ ४४ संचयी-प्रायिकता फलन (Cumulative distribution or distribution function) —

दूसरे अध्याय में संचयी बारंबारता का वर्णन किया जा चुका है। यदि संचयी बारंबारता को कुल बारंबारता से विभाजित किया जाय तो हमें संचयी आपेक्षिक बारंबारता प्राप्त होगी। जिस प्रकार प्रायिकता आपेक्षिक बारंबारता का एक आदर्श स्वरूप है उन्ही प्रकार संचयी आपेक्षिक बारंबारता का आदर्श रूप संचयी प्रायिकता फलन (distribution function) है। इसको $F(x)$ द्वारा सूचित किया जाता है।

$$F(x) = P[X \leq x] \quad (4.5)$$

परंतु यदि सतत चर हो तो

$$P[X \leq x] = \int_{-\infty}^x f(x) dx$$

$$\therefore F(x) = \int_{-\infty}^x f(x) dx \quad (4.6)$$

§ ४४१ संचयी प्रायिकता फलन के गुण

क्योंकि प्रायिकता वक्र और x -निर्देशाक्ष के बीच का कुल क्षेत्रफल 1 होता है, इस कारण $F(x)$ जो इस क्षेत्रफल का वह भाग है जो ऊच्च रेखा $X=x$ के बायी ओर पड़ता है 1 से अधिक नहीं हो सकता। वैसे भी क्यों कि यह X के मान x से कम अथवा उसके बराबर होने तक की प्रायिकता है इसलिए प्रायिकता की भांति इसका मान 0 और 1 के बीच की कोई सख्या ही हो सकता है।

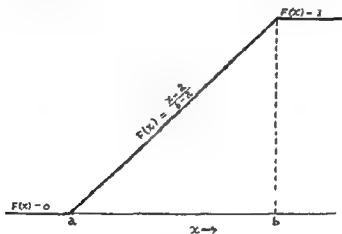
आइए, अब हम देखें कि यदि X का वटन a और b के बीच आयताकार हो तो उसका संचयी प्रायिकता फलन क्या होगा।

$$F(x) = 0 \quad \text{यदि} \quad x \leq a$$

$$F(x) = \frac{x-a}{b-a} \quad \text{यदि} \quad a \leq x \leq b$$

$$F(x) = 1 \quad \text{यदि} \quad x \geq b$$

जैसे दूसरे अध्याय में हमने समष्टि के लिए संचयी बारंबारता चित्र बनाये थे, उसी प्रकार संचयी प्रायिकता फलन को भी चित्र द्वारा निरूपित किया जा सकता है। ऊपर के आयताकार वटन के लिए जो चित्र प्राप्त होगा वह नीचे दिया जा रहा है।



चित्र २०—आयताकार वटन का संचित प्रायिकताफलन

आपका ध्यान इस ओर गया होगा कि इस चित्र में x के बढ़ने के साथ $F(x)$ का मान या तो बढ़ता है या स्थिर रहता है, परंतु कहीं भी x के बढ़ने पर $F(x)$ का मान घटता नहीं। संचयी बारंबारता प्राप्त करने की विधि से ही यह स्पष्ट हो जायगा कि यह बात केवल इस विशेष वटन के लिए ही नहीं बल्कि सभी वटनों के लिए सत्य है।

मान लीजिए कि x_1 और x_2 दो मान हैं जिनमें x_1 छोटा है, यानी $x_1 < x_2$, तो किसी भी वटन के लिए

$$\begin{aligned} F(x_2) &= P(X \leq x_2) \\ &= P[(X \leq x_1) \cup (x_1 < X \leq x_2)] \end{aligned}$$

$$\begin{aligned} &= P(X \leq x_1) + P[x_1 < X \leq x_2] \\ &= F(x_1) + P[x_1 < X \leq x_2] \end{aligned}$$

परंतु क्योंकि $P[x_1 < X \leq x_2]$ का छोटे-से-छोटा मान शून्य ही हो सकता है, इसलिए यदि $x_2 > x_1$ हो तो

$$F(x_2) \geq F(x_1) \quad \dots\dots\dots (47)$$

§ ४५ स्वतंत्र चर (Independent variables) —

तीसरे अध्याय में हम स्वतंत्र घटनाओं की परिभाषा दे चुके हैं। यदि A और B दो स्वतंत्र घटनाएँ हो तो यह सिद्ध किया जा चुका है कि

$$P(A \cap B) = P(A)P(B)$$

यदि (X, Y) एक द्वि विमितीय यादृच्छिक चर हो और हर एक मानमुग्म (a_1, a_2) तथा (b_1, b_2) के लिए

$$\begin{aligned} &P[(a_1 \leq X \leq a_2) \cap (b_1 \leq Y \leq b_2)] \\ &= P[a_1 \leq X \leq a_2] P[b_1 \leq Y \leq b_2] \end{aligned}$$

हो तो यादृच्छिक चर X और Y एक दूसरे से स्वतंत्र कहलाते हैं। इस प्रकार हम देखते हैं कि यदि Y का मान दिया हुआ हो अथवा यह दिया हुआ हो कि Y एक विशेष अंतराल में स्थित है और यदि वह X से स्वतंत्र हो तो दस ज्ञान का X के प्रतिवर्षी प्रायिकता-वंटन पर कुछ भी प्रभाव नहीं पड़ता। इसी प्रकार X के सबंध में किसी प्रतिवर्षी का उससे स्वतंत्र किसी चर Y पर प्रभाव नहीं पड़ता।

यदि X और Y असतत चर हों जो क्रमशः x_1, x_2, \dots, x_m तथा y_1, y_2, \dots, y_n मान धारण कर सकते हों तो

$$P[X=x_i, Y=y_j] = P[X=x_i] P[Y=y_j] \quad (48)$$

$i=1, 2, \dots, m, \quad j=1, 2, \dots, n$

इस प्रकार यदि हमें X और Y के वंटन ज्ञात हो और यदि यह भी मालूम हो कि ये दोनों चर स्वतंत्र हैं तो हम इन दोनों का संयुक्त-वंटन (Joint distribution) इनके अलग-अलग वंटनों के गुणन से प्राप्त कर सकते हैं।

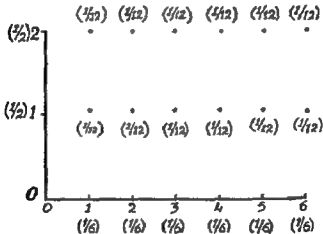
इसी प्रकार यदि सतत चर X और Y स्वतंत्र हों, उनके घनत्व फलन क्रमशः $f_1(x)$ तथा $f_2(y)$ हों, और उनके संयुक्त वंटन का घनत्व-फलन $f(x, y)$ हो तो

$$f(x, y) = f_1(x) f_2(y) \quad \dots \quad (49)$$

सम्युक्त-वटन के घनत्व फलन की परिभाषा भी उसी प्रकार दी जा सकती है जिस प्रकार किना एक विमितीय यादृच्छिक चर के घनत्व फलन की

$$f(x, y) = \frac{I}{\delta_1 \delta_2} \frac{P[(x - \delta_1 < X < x + \delta_1) \cap (y - \delta_2 < Y < y + \delta_2)]}{(\delta_1 + \delta_1) (\delta_2 + \delta_2)}$$

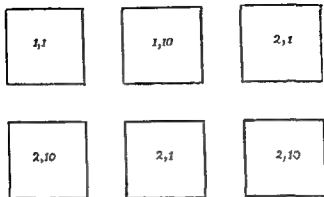
उदाहरण (१)—एक पाँसा और एक रुपया साथ-साथ उछाले जाते हैं। X एक यादृच्छिक चर है जिसका मान पाँसे के ऊपर के मुख पर प्राप्त बिंदुओं के बराबर है। Y भी एक यादृच्छिक चर है। यदि रुपया चित पड़े तो इसका मान 1 होता है, यदि वह पट पड़े तो इसका मान 2 होता है। ये दोनों यादृच्छिक चर स्पष्ट तथा स्वतंत्र हैं, इसलिए इनका सम्युक्त वटन नीचे दिये हुए चित्र के अनुसार होगा।



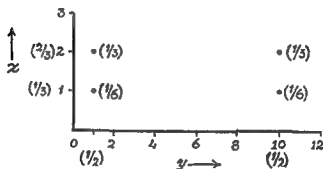
चित्र २१—दो स्वतंत्र यादृच्छिक चरों के सम्युक्त और एक-पारिभाषित घटन

(२) अब मान लीजिए कि एक पाँसे के प्रत्येक मुख पर बिंदुओं के स्थान पर दो-दो समस्याएँ लिखी हुई हैं जो नीचे दिये हुए चित्र के अनुसार हैं।

पाँसे को फेंकने से जो मुख ऊपर की ओर आता है उस पर लिखी हुई पहिली समस्या को ξ और दूसरी समस्या को μ से सूचित किया जाय तो ξ और η का सम्युक्त-वटन चित्र 22 के अनुसार होगा।



चित्र २२—एक पसि के छ. मूल



चित्र २३—चित्र २२ में दशित पसि को फेंकने से प्राप्त ऊपर की संख्याओं का समुक्त वंटन

इस उदाहरण से हमें यह मालूम पडता है कि दो यादृच्छिक चरों में भौतिक संबध होते हुए भी वे एक दूसरे से स्वतंत्र हो सकते हैं ।

§ ४६ प्रायिकता वंटन के प्रति समाकलन (Integration with respect to a probability distribution)

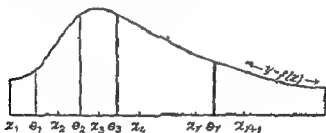
मान लीजिए कि X एक असतत यादृच्छिक चर है जो a_1, a_2, \dots, a_n आदि n मान धारण करता है ।

मान लीजिए $g(X)$ यादृच्छिक चर X का एक फलन है और $P(x) = P(X=x)$ तब

$$\sum g(x) P(x) = g(a_1) P(a_1) + g(a_2) P(a_2) + \dots + g(a_n) P(a_n)$$
 को हम X के प्रायिकता वटन के प्रति समाकलन कहते हैं और इस समाकलन को $\int g(x) dF(x)$ से सूचित करते हैं।

$$\begin{aligned} dF(x) &= F(x) - F(x-dx) \\ &= P[x-dx < X < x] \\ &= P(x) \end{aligned}$$

यदि dx इतना छोटा हो कि $x-dx$ और x के बीच में X का कोई भी सम्भव मान a_1, a_2 आदि न हों। यदि $P(x)$ के स्थान पर हम समष्टि की आपेक्षिक बारंबारता को रखें तो हम देख सकते हैं कि हमें इस प्रकार $g(X)$ का औसत मान प्राप्त हो जायगा। इसी प्रकार आपेक्षिक बारंबारता के स्थान पर उसके आवर्त रूप प्रायिकता के होने पर यह समाकलन $g(X)$ का प्रत्याशित मान अथवा माध्य देता है।



चित्र २४

यदि यादृच्छिक चर सतत है और उसका घनत्व-फलन $f(x)$ हो तो इस चर के परास को छोटे-छोटे भागों में विभाजित किया जा सकता है। मान लीजिए, इस प्रकार के विभाजन की क्रम सख्या दी हुई है और r वें भाग में X का एक मान θ_r है। तब हम एक योग का कलन कर सकते हैं जो निम्नलिखित है—

$$\sum g(\theta_r) f(\theta_r) (x_{r+1} - x_r)$$

जहाँ x_r और x_{r+1} उस अंतराल के सीमान्त बिन्दु हैं जिसमें θ_r स्थित है। यदि हम इन विभाजनों को छोटा करते चले जायें और इस प्रकार उनकी सख्या बढ़ाते चले जायें तो यह योग एक निश्चय मान की ओर अभिसर होता है। जिस मान की ओर यह योग अभिसर होता है उसे हम X के प्रायिकता वटन के प्रति $g(x)$ का समाकलन कहते

है। इस समाकलन को हम $\int_{-\infty}^{\infty} g(x) f(x) dx$ द्वारा सूचित करते हैं। क्योंकि x पर प्रायिकता घनत्व $= f(x)$, इसलिए $x-dx$ और x के बीच का प्रायिकता द्रव्यमान $= f(x) dx$

$$= F(x) - F(x-dx)$$

$$= dF(x)$$

$\int g(x) dF(x)$ एक ऐसा संकेत है जो हम दोनों प्रकार के चरों—सतत और असतत के लिए प्रयोग कर सकते हैं। इस प्रकार

$$\int g(x) dF(x) = \sum g(a_i) P(a_i) \text{ यदि } X \text{ असतत हो}$$

$$\text{तथा } \int g(x) dF(x) = \int_{-\infty}^{\infty} g(x) f(x) dx \text{ यदि } X \text{ सतत हो।}$$

§ ४.७ यादृच्छिक चर का प्रत्याशित मान अथवा माध्य (*Expected value or mean value of a random variable*)—

मान लीजिए कि $g(X) = X$ तब $\int x dF(x)$ को हम यादृच्छिक चर X का माध्य अथवा प्रत्याशित मान कहते हैं। और इसे $E(X)$ से सूचित करते हैं। यह आपको याद होगा कि यदि आंकड़ें आवृत्ति सारणी में दे रखे हों तो माध्य के लिए निम्नलिखित सूत्र का उपयोग होता है।

$$\bar{x} = \frac{\sum_{i=1}^n x_i \frac{f_i}{n}}{\sum_{i=1}^n \frac{f_i}{n}}$$

यदि X एक असतत चर है तो

$$E(X) = \sum_{i=1}^n x_i P(x_i)$$

इसी प्रकार X के किसी फलन $g(X)$ का प्रत्याशित मान

$$E[g(X)] = \int g(x) dF(x)$$

इन दोनों सूत्रों में बहुत अधिक समानता है। यदि आपेक्षिक बारबारता $\frac{f_i}{n}$

$$\sum_{i=1}^n \frac{f_i}{n}$$

की जगह हम प्रायिकता $P(x_i)$ को रखें जो वास्तव में इस आपेक्षिक बारंबारता का आदर्श रूप है तो हमें यादृच्छिक चर का माध्य प्राप्त हो जाता है। इन दोनों में विशेष अंतर यही है कि पहले सूत्र का प्रयोग समष्टि पर किया जाता है जिसके बारे में हमें पूर्ण ज्ञान है, परन्तु दूसरे सूत्र का प्रयोग यादृच्छिक चर के लिए किया जाता है। यादृच्छिक चर किसी विशेष प्रयोग में क्या मान धारण करेगा यह अनिश्चित रहता है। अतः हमें प्रायिकता के शब्दों में ही बात करनी पड़ती है।

§ ४८ यादृच्छिक चर के घूर्ण (Moments of a random variable)

जिस प्रकार समष्टि में मध्यांतरित r वां घूर्ण

$$\mu_r = \frac{\sum_{i=1}^n (x_i - m)^r \cdot \frac{f_i}{n}}{\sum_{i=1}^n \frac{f_i}{n}}$$

होता है उसी प्रकार यादृच्छिक चर का r वां घूर्ण $\mu_r = \int [x - E(X)]^r dF(x)$ होता है। इसके दूसरे मध्यांतरित घूर्ण $\mu_2 = \int [x - E(x)]^2 dF(x)$ को चर का प्रसरण (variance) कहते हैं। अधिकतर $E(x)$ को μ तथा $E(X - \mu)^2$ को $V(X)$ से सूचित किया जाता है। समष्टि के चर की भाँति ही यादृच्छिक चर के a -आंतरिक घूर्णों की परिभाषा भी दी जा सकती है। a -आंतरिक और मध्यांतरित घूर्णों का एक दूसरे से संबंध भी उसी प्रकार का होता है।

§ ४९ स्वतंत्र चरों के गुणनफल का प्रत्याक्षित मान

यदि बटन असतत हों तो

$$\begin{aligned} E(XY) &= \sum_{i=1}^m \sum_{j=1}^n x_i y_j \cdot P[X=x_i, Y=y_j] && \text{[दिए गए]} \\ &= \sum_{i=1}^m \sum_{j=1}^n x_i y_j \cdot P[X=x_i] P[Y=y_j] && \text{समीकरण (4.8)} \\ &= \sum_{i=1}^m x_i P[X=x_i] \sum_{j=1}^n y_j P[Y=y_j] \\ &= E(X)E(Y) && (4.10) \end{aligned}$$

यह सूत्र सतत बटनों के लिए भी आसानी से सिद्ध किया जा सकता है

§ ४१० चरों के योग का प्रत्याशित मान

$$\begin{aligned}
 E(X+Y) &= \sum_{i=1}^m \sum_{j=1}^n (x_i + y_j) P(X=x_i, Y=y_j) \\
 &= \sum_{i=1}^m \sum_{j=1}^n x_i P(X=x_i, Y=y_j) + \sum_{i=1}^m \sum_{j=1}^n y_j P(X=x_i, Y=y_j) \\
 &= E(X) + E(Y) \qquad (417)
 \end{aligned}$$

यह सूत्र सतत घटनों के लिए भी सरलता से सिद्ध हो सकता है।

भाग २

परिकल्पना की जाँच (*Testing of Hypothesis*)

और

कुछ महत्वपूर्ण प्रायिकता वंटन (*Probability Distributions*)

अध्याय ५

मनोवैज्ञानिक पृष्ठभूमि

§ ५ १ क्या वचन में आपको परियों की कहानी पढ़ने का शौक रहा है ? यदि हाँ तो आपने उस विचित्र वर्तन के बारे में अवश्य सुना होगा जिसमें शहद भरा रहता था और चाहे जितना शहद उसमें से निकाल लें वह खाली नहीं होता था । यदि मैं आपको शहद से भरा हुआ एक बर्तन देकर कहूँ कि लीजिए यही वह प्रसिद्ध बर्तन है जिनके बारे में आपने वचन में बहुत कुछ पढ़ा-सुना होगा तो आप मेरे इस कथन की जाँच कैसे करेंगे ?

आप कहेंगे कि इस कथन की सच्चाई की जाँच करने में क्या रखा है । अपने मित्रों को एक पार्टी दीजिए और उसमें सबको काफी मात्रा में शहद बाँट दीजिए । यदि बर्तन खाली हो जाता है तो कथन गलत है । लेकिन कल्पना कीजिए कि बर्तन वास्तव में भरा का भरा ही रहता है तब आपको आश्चर्य होगा और कदाचित् मेरे कथन की सच्चाई में विश्वास ही हो जाय । लेकिन यदि आपका दृष्टिकोण आलोचनात्मक है तो आप निश्चय ही मेरे कथन को सत्य मानना पसन्द नहीं करेंगे । आप कह सकते हैं कि यद्यपि इस प्रथम जाँच में यह बर्तन खाली नहीं हुआ, परन्तु इससे यह तो सिद्ध नहीं होता कि यह वही बर्तन है जिसका कहानियों में वर्णन है । वह तो कभी खाली होता ही नहीं था । यदि यह बर्तन प्रथम प्रयास में खाली नहीं हुआ तो यह नहीं कहा जा सकता कि यह कभी खाली होगा ही नहीं । फिर भी यदि बर्तन बार-बार जाँच में उत्तीर्ण हो तो आपका विश्वास मेरे कथन पर दृढ़तर होता जायगा ।

§ ५ २ इस प्रकार हम देखते हैं कि यदि किसी कथन से ऐसा निष्कर्ष निकलता है जो अनुभव में विपरीत है तो हम उस कथन को झूठ समझते हैं । परन्तु यदि अनुभव में निष्कर्ष के अनुकूल है तब भी हम यह नहीं समझ बैठते कि कथन सिद्ध हो गया । बल्कि केवल उस कथन में हमारा विश्वास दृढ़तर होता जाता है । यदि आपको परियों की कहानियों में न तो दिलचस्पी हो और न विश्वास तो उस दशा में आप उपयुक्त कथन के प्रयोग करने का भी कष्ट न करेंगे और प्रारम्भ से ही मुझे झूठा समझेंगे ।

यद्यपि बिना प्रयोग के ही अपना मत स्थिर कर लेना किसी वैज्ञानिक के लिए उचित नहीं है, फिर भी आपने इस मत से मुझे कुछ विरोध नहीं है। इसके लिए एक विश्वसनीय उदाहरण देता हूँ।

§ ५३ श्रामुत 'क' पर आरोप लगाया जाता है कि उन्होंने 'ख' का खून किया है। यह कहा जा सकता है कि २५ सितम्बर की रात को श्री 'ख' कलकत्ते से दिल्ली जानेवाली गाड़ी में बहुत-सा धन लेकर यात्रा कर रहे थे। श्री 'क' उनके डिब्बे में घुस गये और श्री 'ख' के सो जाने पर उन्होंने धन चुराने का प्रयास किया। परन्तु श्री 'ख' को अचानक नींद टूट जाने पर उन्होंने शोर-गुल मचाना चाहा। यह देखकर श्री 'क' घबरा गये उन्होंने पिस्तौल निकालकर उसी दम श्री 'ख' का काम तमाम कर दिया।

यह पुलिस का कहना है। पुलिस ने श्री 'क' को तीन दिन पश्चात् दिल्ली में गिरफ्तार किया जब उनके पास उन नोटों में से कुछ पाये गये जो श्री 'ख' के पास दिल्ली जाते समय थे। आइये, जिस सिद्धान्त का प्रतिपादन हमने परियों की कहानी में किया था उसका प्रयोग पुलिस के इस कथन पर करके देखें।

कथन है "श्री 'क' ने श्री 'ख' को २५ सितम्बर की रात में कलकत्ते से जानेवाली रेलगाड़ी में मार डाला।"

यदि यह कथन सच है तो यह निष्कर्ष निकलता है कि २५ सितम्बर की रात को 'क' और 'ख' एक ही रेलगाड़ी में यात्रा कर रहे थे। यदि यह निष्कर्ष गलत सिद्ध हो जाय तो उपर्युक्त कथन भी स्वभावतः गलत सिद्ध हो जायगा। मान लीजिए कि कई गवाह शपथपूर्वक यह कहने को तैयार हैं कि 'क' २५ सितम्बर की रात को दिल्ली में थे और यही नहीं २४ तारीख से ही वे दिल्ली में रह रहे हैं। इस गवाही के बाद और यह जानते हुए कि एक ही व्यक्ति एक ही समय पर दो विभिन्न स्थानों में नहीं रह सकता, मूल कथन झूठा सिद्ध हो जाता है।

इसके विपरीत मान लीजिए कि कुछ गवाह इस निष्कर्ष की पुष्टि करते हैं कि श्री 'क' और श्री 'ख' एक ही रेलगाड़ी से यात्रा कर रहे थे। इस गवाही से यह सिद्ध नहीं होता कि 'क' ने 'ख' का खून किया था। परन्तु पुलिस का कथन इस कारण अधिक विश्वसनीय हो जाता है।

यदि पुलिस के कथन से अनेकों निष्कर्ष निकाले जायें जिनकी पुष्टि गवाहों द्वारा हो तो न्यायाधीश का विश्वास उनकी कहानी की सच्चाई में क्रमशः दृढतर होकर प्रायः असंदिग्धता में परिणत हो सकता है। फिर भी निष्कर्ष के प्रतिकूल एक

भी गवाही मिलने पर उन सब गवाहियों का प्रभाव नष्ट हो जाता है जो कथन के निष्कर्षों के अनुकूल थी।

मान लीजिए कि निम्नलिखित बातें सिद्ध हो जाती हैं—

(१) 'क' 'ख' से परिचित था।

(२) 'ख' के खून के कुछ ही दिन पूर्व 'क' और 'ख' में किसी जमीन के टुकड़े के स्वामित्व को लेकर बहुत झगडा हुआ था।

(३) 'क' और 'ख' एक ही गाडी से यात्रा कर रहे थे।

(४) जब 'क' दिल्ली से रवाना हुआ तब उसके पास प्रायः कुछ भी नहीं था। परन्तु जब वह पकडा गया तो उसके पास लगभग १,००० रुपया निकला।

जब बादी उगत घटनाओं की पुष्टि गवाही द्वारा कर चुका हो तो एक और घटना प्रकाश में आती है—

(५) जब 'ख' ने दिल्ली के लिए टिकट खरीदा तो 'क' ने उसका पीछा किया और उसी डिब्बे में एक सीट रिज़र्व करा ली।

यदि घटना नम्बर (३) पहिले ही ज्ञात नहीं होती तो इस नयी घटना से बादी के कथन की सचाई में विश्वास बहुत बढ जाता। परन्तु घटना नम्बर (३) के सिद्ध होने के पश्चात् इसका महत्त्व पहिले की अपेक्षा बहुत कम हो जाता है। फिर भी यदि हम घटना नम्बर (४) पर विचार करें तो घटना नम्बर (३) के सिद्ध होने के पश्चात् भी इससे बादी के कथन को काफी बल मिलता है।

§ ५४ यदि नवीन साक्ष्य विश्वसनीय पूर्वज्ञात घटनाओं से बहुत अधिक संबंधित हो तो साक्ष्य में हमें अधिक विश्वास होगा। परन्तु इस साक्ष्य से हमारे विश्वासों में अन्तर नहीं पडता। और यदि पडता भी है तो अधिक नहीं। इसके विपरीत यदि यह नवीन साक्ष्य पूर्व ज्ञात घटनाओं से एकदम असंबंधित हो तो यह हमारे पूर्व निश्चित विचारों को बल देने में अथवा उनको खंडित करने में बहुत महत्त्व रखता है।

मनुष्य का मस्तिष्क प्रायः इसी प्रकार कार्य करता है। यह ऐसा क्यों करता है? यह ऐसा प्रश्न है जिसकी इस पुस्तक में चर्चा करना उचित प्रतीत नहीं होता। इस कार्य के लिए कदाचित् कोई मनोवैज्ञानिक ही सबसे अधिक उपयुक्त है। बल्कि हमें विश्वास है कि उसे भी इसका उत्तर देने में बहुत कठिनाई पडेगी। संभवतः उसका मस्तिष्क भी इसी प्रकार कार्य करता है और वह हमें इस समस्या के अपने हल के बारे में विश्वास दिलाने के लिए जो युक्तियाँ देगा उसमें भी वह इस सिद्धान्त का प्रयोग करेगा। इसके अलावा हम इस बात की भी चर्चा नहीं करेंगे कि इन सिद्धान्तों का प्रयोग

कहाँ तक युक्तियुक्त है। यह असंभव है कि इस प्रकार का कोई भी तर्क मूढ़ और जटिल न हो जाये। विभिन्न व्यक्तियों की भिन्न-भिन्न राय हो सकती है। सबसे कठिन समस्या तो यह निश्चय करने की है कि युक्तियुक्त आचरण क्या है। साहित्यकी की एक पुस्तक का लेखक, जो अपने परिहासशील स्वभाव के लिए जरा भी प्रसिद्ध नहीं है तथा जो एक गम्भीर वैज्ञानिक माना जाता है, युक्तियुक्त आचरण की परिभाषा देते हुए लिखता है कि यह वह आचरण है जिसे वह लेखक युक्तियुक्त समझता है। यद्यपि इस प्रकार की कोई भी परिभाषा बिलकुल भी युक्तिमग्न ज्ञात नहीं होती तथापि यह हो सकता है कि पाठकों का बहुमत इस लेखक के साथ हो। इस परिभाषा के बारे में ही नहीं किन्तु इस बारे में भी कि निर्णय किस प्रकार किया जाये और निष्कर्ष कैसे निकाला जाये।

§ ५५ हमने ऊपर यह दिखलाया है कि मानव मस्तिष्क किसी कथन के अनुमोदन में अथवा उसके विपरीत साक्ष्य को किस प्रकार तौलता है। प्रायः ऐसी ही बात उस समय भी दृष्टिगोचर होती है जब कथन का निष्कर्ष झूठ या गलत तो नहीं सिद्ध होता, परन्तु निष्कर्ष असंभाव्य (improbable) मालूम होता है। कई लोगों का, जो सिनेमा को बहुत आलोचनात्मक दृष्टिकोण से देखते हैं, यह मत है कि भारतीय चित्रों में क्या, घटना-चक्र, काल और वातावरण बनावटी तथा वास्तविकता से बहुत दूर होता है। मनुष्यों का जो आचरण और व्यवहार उसमें दिखाया जाता है वह प्रायः अस्वाभाविक होता है। उदाहरण के लिए अभिनेता का कोढ़े द्वारा पीटे जाने और नयकर पीडा दिये जाने पर माना अथवा अभिनेत्री का अपनी माँ की मृत्यु पर आंसू बहाने के साथ साथ गीत गाना। स्त्रियों को ऐसे स्वन पहने हुए दिखाया जाता है कि जो पहले कभी नहीं देखे गये यद्यपि चित्र के पश्चात् उनका काफी चलन हो सकता है। एक पढ़े लिखे सम्मानित व्यक्ति को सड़को पर नाचता और गाता हुआ दिखाया जाता है। इन सभी दशाओं में आलोचनात्मक दृष्टिकोणवाले व्यक्तियों का यह विचार होता है कि यह सब बनावटी और अस्वाभाविक है। जब कोई यह कहता है कि कोई आचरण या घटना अस्वाभाविक है तब इसके अर्थ यहाँ होते हैं कि साधारण-तया कोई मनुष्य इस तरह की घटनाओं की अथवा आचरण की आशा नहीं करता। यदि चित्र में ये दिखाये जाते हैं तो आपके मन में बराबर यही विचार आयेगा कि वास्तविक जीवन में ऐसा कभी नहीं हो सकता। यहाँ तक कि यदि निर्माता चित्र के आरम्भ में यह धोषणा भी कर दे कि चित्र के पात्र और घटनाएँ वास्तविक जीवन से ही ली गयी हैं तब भी आपको विश्वास नहीं होगा।

आखिर ऐसा नयो ? क्या यह असंभव है कि कोई लडकी अपनी माँ के मरने पर एक दुःख भरा गीत गाये ? मुझे तो यह असंभव नहीं भात्रूम पडता यद्यपि किसी भी लडकी से इस प्रकार के आचरण की कोई भी आशा नहीं रखता । दूसरे इस प्रकार के आचरण की संभावना भी बहुत कम है । यदि आप इसे प्रायिकता की भाषा में व्यक्त करना चाहें तो कह सकते हैं कि इस घटना की प्रायिकता बहुत कम है । यद्यपि इस प्रायिकता का ठीक-ठीक मान मयबा अनुमान किसी को भी नहीं मालूम होता । लेकिन यदि हम यह कहें कि प्रायिकता इस सहूल में एक से कम है तो कदाचित् भूल नहीं होगी । जब हमें कोई कभी ऐसी घटना का घर्षण सुनाता है जिसकी प्रायिकता बहुत कम हो तो उस पर हमें सहज ही विश्वास नहीं हो जाता ।

मान लीजिए कि कोई व्यक्ति एक ऊँचे मकान की छत से सडक पर कूद पडता है । साधारणतया हम यह अपेक्षा करते हैं कि यदि वह व्यक्ति मरने से बच भी गया तो घुरी तरह आहत तो अवश्य ही होगा । यदि किसी चित्र में यह दिखाया जाय कि एक लडका इस प्रकार कूदता है और आहिस्ता से सडक पर जाकर भीड में मिल जाता है जहाँ कोई इस बात पर ध्यान भी नहीं देता तो कदाचित् दर्शकों का इस वृष्य से मनोविनोद तो अवश्य होगा, परन्तु कोई भी गभीरतापूर्वक ऐसी घटना के वास्तविक जीवन में घटने की कल्पना नहीं कर सकेगा ।

अब यही घटना यदि सिनेमा में नहीं दिखाई जाती बल्कि एक ऐसे व्यक्ति द्वारा आपको सुनाई जाती जिसकी ईमानदारी में आपको पूरा भरोसा है और यदि वह यह कहता कि उसने यह घटना स्वयं देखी है तो आप पर इसका क्या प्रभाव पडता ? शायद शुरू में आप सोपते कि उस व्यक्ति को कुछ धोखा हुआ हो, परन्तु यदि वह बहुत बलपूर्वक अपने कथन का समर्थन करे और उसके मस्तिष्क के संतुलन और वैज्ञानिक प्रेक्षण की आदत से आप परिचित हों तो आपको उसकी बात का विश्वास करना होगा । आपको आश्चर्य तो अवश्य होगा परन्तु आप यही सोचेंगे कि एक बहुत ही विचित्र घटना घटी ।

क्या कारण है कि एक ही घटना के बिलकुल एक ही प्रकार के शब्दों के दो भिन्न व्यक्तियों द्वारा बिये गये वर्णनों की इतनी विभिन्न प्रतिक्रिया होती है ? पहले व्यक्ति के बारे में आप जानते हैं कि उसे विचित्र बातें गड कर सुनाने का शौक है या झूठ बोलने में उसे कोई हिचकिचाहट नहीं होती । इस दशा में यदि वह किसी अनहोनी घटना का वर्णन करता है तब आप यही समझते हैं कि यह गप्प लगा रहा है । दूसरे व्यक्ति के बारे में आप यह जानते हैं कि वह अपने जीवन में आज तब झूठ बोलता ही नहीं । ऐसी

दशा में आपको यह समझ न मालूम होगा कि आज वह बिना कारण आपसे झूठ बोल रहा है। अब दो घटनाएँ हैं और दोनों ही की प्रायिकताएँ बहुत कम हैं। एक तो यह घटना है कि एक लड़का घर की तीसरी मजिल से भरी हुई सबक पर बिना किसी दुर्घटना के और बिना किसी का ध्यान आकर्षित किये कूद जाता है और दूसरी घटना यह है कि एक मनुष्य जिसने आज तक झूठ नहीं बोला आज बिना कारण झूठ बोल रहा है। यदि इन दो घटनाओं की प्रायिकता की तुलना करने पर—यद्यपि हमारे पास इन प्रायिकताओं का सही मान प्राप्त करने का कोई तरीका नहीं है परन्तु केवल अवचेतन मन में ही यह तुलना समझ है—आप यह तय करते हैं कि उस मनुष्य को झूठ बोलने की संभावना इस अनहोनी घटना से भी कम है तब आपको उस मनुष्य का विश्वास हो जायेगा, और आप यही सोचेंगे कि कौमी विचित्र घटनाएँ घट सकती हैं !

इस सारे विवाद का तात्पर्य यह है कि ऐसी घटनाओं में किसी को सहज ही विश्वास नहीं होना जिनकी प्रायिकता बहुत कम होनी है। यदि किसी कथन से कुछ ऐसा निष्कर्ष निकलना हो जिसके होने की संभावना बहुत कम हो तो पहिले तो हम यह तय करते हैं कि निष्कर्ष सत्य नहीं हो सकता, क्योंकि इसकी प्रायिकता बहुत कम है। इस निष्कर्ष को असत्य मानने का स्वाभाविक परिणाम होता है कि हम उस कथन को भी असत्य मान लेंगे है जिससे इस विचित्र और अविश्वसनीय निष्कर्ष का जन्म हुआ था।

§ ५६ यही वह मनोवैज्ञानिक पृष्ठभूमि है जिस पर परिकल्पना की जाँच का सांख्यिकीय सिद्धान्त (Statistical theory of testing of hypothesis) आधारित है। इस प्रकार के मनोवैज्ञानिक आचरण को जो एक साधारण मनुष्य के लिए स्वाभाविक है और जिसके लिए वह किसी प्रकार के सोचने-विचारने की आवश्यकता नहीं समझता, सांख्यिकी के विद्वानों ने तर्क द्वारा युक्ति-संगत ठहराया है। मान लीजिए कि उन सब घटनाओं को जिनकी प्रायिकता एक प्रतिशत या उससे कम हो इन असंभव समझ लें और ऐसी घटनाओं से संबंधित कथन को झूटा या गलत समझें तो हमारे इस निष्कर्ष के गलत होने की प्रायिकता भी एक प्रतिशत से कम ही होगी। यदि कथन वास्तव में झूठा है तो हमारा निष्कर्ष सत्य ही है। और यदि कथन सत्य है तो हम उस निष्कर्ष को उस समय ही झूठा मानेंगे जब वह असंभव मालूम होनेवाली घटना सचमुच घटित हो जाय। क्योंकि हम जानते हैं कि उस घटना की प्रायिकता एक प्रतिशत से कम है, इसलिए इस प्रकार उपर्युक्त सिद्धान्त के अनुसार कथनों को झूठा मानने की प्रायिकता भी एक प्रतिशत से कम ही होगी। विद्वानों के इस दृष्टिकोण का विकास हम अगले अध्यायों में करेंगे जिसमें कुछ विस्तार से इस प्रकार की युक्ति

और दर्शन पर विचार होगा। यहाँ तो हम केवल सांख्यिकीय पद्धति से जाँच के कुछ उदाहरण देंगे और ऐसे प्रायिकता वटनों का परिचय करायेंगे जो बहुत महत्वपूर्ण और उपयोगी हैं।

§ ५७ मान लीजिए कि एक रोग है जिससे पीड़ित अधिकतर रोगी मृत्यु का शिकार हो जाते हैं। वैज्ञानिक अवश्य ही ऐसे रोग के इलाज के लिए औषध की खोज में सलग्न होंगे। उनको यह पता है कि—

(१) इस रोग से पीड़ित सभी व्यक्ति नहीं मर जाते। कुछ ठीक भी हो जाते हैं।

(२) किसी भी औषध से सब रोगी ठीक नहीं हो जाते।

(३) यद्यपि किसी विशेष औषध से वह विशेष रोग ठीक हो जाये जिसके लिए वह औषध दी गयी थी तथापि यह संभव है कि रोगी को अन्य कोई रोग भी हो और औषध का ठीक प्रभाव होते हुए भी वह मर जाये।

इस दशा में यदि उस औषध के उपयोग में मृतकों के अनुपात में कमी हो सके और वह पुराने उच्च स्तर से गीचे उतर आये तो यह सचमुच ही प्रगति का सूचक है। नयीन औषध का उपयोग वास्तव में ठीक दिशा में प्रभाव डाल रहा है अथवा नहीं यह निर्णय करने के लिए यह जानने की आवश्यकता है कि किस समय कोई औषध नहीं दी जाती थी उस समय रोगियों में मरनेवालों का अनुपात क्या था तथा इस औषध के देने से इस अनुपात में क्या अन्तर पड़ा।

कारणों कीजिए कि सैंकड़ों डाक्टरों के अनुभव के आधार पर, जिन्होंने इस रोग से पीड़ित हजारों व्यक्तियों को देखा है, हमें यह ज्ञात है कि इस प्रकार के रोगियों में मृतक-अनुपात २०% है। अब जिस नयी औषध से इस रोग के इलाज में प्रगति की आशा की जाती है उसका प्रयोग हम अनियमित अथवा यादृच्छिक रूप से चुने हुए सौ रोगियों पर करते हैं। यदि हमारा प्रतिदर्श (sample) कुल रोगियों का सच्चा प्रतिनिधि है—उदाहरण के लिए रोगियों की उम्र और उनके रोग की दशा कुल रोगियों के समूह में और इस प्रतिदर्श में समान अनुपात में है—और यदि इस नयी औषध से कुछ लाभ नहीं होता तो इन सौ रोगियों में से २० की मृत्यु की आशंका है। या तो २० की ही मृत्यु होगी या संयोग से कुछ कम या अधिक व्यक्ति भी मर सकते हैं। यदि यह मान लिया जाये कि औषध का प्रभाव रोग पर कुछ भी नहीं होता तो रोगियों में से केवल दस मरने की प्रायिकता किन्तनी है?

यदि यह प्रायिकता इतनी काफी है कि संयोग से ऐसी घटना होने पर हमें कुछ भी आश्चर्य नहीं होगा तो हम यही कह सकते हैं कि कदाचित् इस औषध का कुछ गुण-

कारी प्रभाव इस रोग पर पड़ता हो, परन्तु इस प्रयोग से जो एक सौ रोगियों पर किया गया यह दावा सिद्ध नहीं होता। इसके बारे में अधिक निश्चित होने के लिए हमें प्रतिदरशों को और भी बड़ा करने की आवश्यकता है। इस प्रकार की मनोवैज्ञानिक प्रतिक्रिया की हम आशा रखते हैं क्योंकि यह कथन कि इस औपध से कुछ लाभ नहीं होता उसी समय झूठा माना जायगा जब कि प्रेक्षित मृत्यु-संख्या की प्रायिकता ऊपर लिखी हुई परिकल्पना के आधार पर बहुत ही कम निकले। यदि यह प्रायिकता काफी बड़ी हो तो कोई कारण नहीं है कि इस परिकल्पना को झूठा माना जाये। फिर भी यदि प्रेक्षित मृत्यु-संख्या उस संख्या में कम है जिसकी आशंका थी तो हो सकता है कि वास्तव में औपध गुणकारी हो। परन्तु निश्चयपूर्वक जानने के लिए और अधिक प्रेक्षणों की आवश्यकता है।

इसके पूर्व कि हम यह कह सकें कि क्या संख्या प्रायः सभ्य है और क्या नहीं, हमें यह ज्ञात होना चाहिए कि प्रायिकता की गणना कैसे की जाये। भिन्न-भिन्न मृत्यु-संख्याओं की प्रायिकता हमें मालूम होनी चाहिए। यदि चिकित्सा से कुछ लाभ नहीं होता तो रोगियों में मृत्यु को प्राप्त होनेवालों का अनुपात २०% होना चाहिए। भिन्न-भिन्न संख्या के प्रतिदरशों में इस अनुपात में कहीं तक अंतर पड़ सकता है ?

यदि हम केवल एक रोगी पर प्रयोग करके देखते हैं तो दो घटनाओं की संभावना है, या तो वह ठीक हो जायेगा या उसकी मृत्यु हो जायेगी। पहली दशा में प्रतिदरशों में मृतको का अनुपात शून्य प्रतिदरश है जब कि दूसरी दशा में यह अनुपात शत प्रतिदरश होगा। पहली दशा में यह अनुभव से प्राप्त औसत अनुपात से (जो २०% है) बहुत कम होगा। परन्तु यह इस बात का कोई प्रमाण नहीं है कि औपध वास्तव में गुणकारी है। बिना इस औपध के भी ८०% लोग ठीक हो ही जाते थे और यदि यह विदोष रोगी ठीक हो जाता है तो इसमें आश्चर्य की कोई बात नहीं। इसी प्रकार रोगी के मरने पर यह कहना भी ठीक नहीं कि इस औपध से कुछ भी लाभ नहीं होता या इससे हानि ही होती है। इस प्रकार यह मालूम होता है कि केवल एक रोगी पर प्रयोग करके हम किसी निश्चित मत पर नहीं पहुँच सकते। इसने लिए हमें अधिक रोगियों पर परीक्षण करना आवश्यक है।

अब यदि दो रोगियों पर प्रयोग किया जाये तो निम्न तीन घटनाओं की संभावना है—

- (१) दोनों रोगी मर जायें।
- (२) एक रोगी मर जाये और एक ठीक हो जाये।

(३) दोनों रोगी ठीक हो जायें ।

यदि औषध का कुछ प्रभाव न हो तो एक रोगी के मरने की प्रायिकता $P(A) = \frac{100}{100} = \frac{1}{100}$ है और उसके ठीक हो जाने की प्रायिकता $P(B) = \frac{99}{100} = \frac{99}{100}$ है । इसी प्रकार दूसरे रोगी के मरने की प्रायिकता भी $\frac{1}{100}$ है । यह युक्तिसंगत माना जा सकता है कि एक रोगी की मृत्यु का दूसरे रोगी के ठीक होने से या उसकी मृत्यु होने से कुछ भी संबंध नहीं है । अर्थात् ये दोनों घटनाएँ स्वतंत्र हैं । इस कारण दोनों रोगियों के मरने की प्रायिकता

$$\begin{aligned} P(A \cap B) &= P(A)P(B) \\ &= \frac{1}{100} \times \frac{1}{100} \\ &= \frac{1}{10000} \end{aligned}$$

यदि रोगियों को 'क' और 'ख' से सूचित किया जाये तो इस घटना की प्रायिकता कि 'क' मर जाये और 'ख' ठीक हो जाये $\frac{1}{100} \times \frac{99}{100} = \frac{99}{10000}$ है । इसी प्रकार 'क' के ठीक हो जाने और 'ख' के मरने की प्रायिकता $\frac{99}{100} \times \frac{1}{100} = \frac{99}{10000}$ है । इस दूसरी घटना— कि एक रोगी मर जाये और एक ठीक हो जाये—की प्रायिकता ऊपर लिखी दोनों अपवर्जी घटनाओं (exclusive events) की प्रायिकताओं के योग से प्राप्त होगी । अर्थात् इस घटना की प्रायिकता $\frac{99}{5000}$ है ।

दोनों रोगियों के ठीक हो जाने की प्रायिकता $\frac{99}{100} \times \frac{99}{100} = \frac{9801}{10000}$ है । हम इन प्रायिकताओं को एक सारणी के रूप में निम्न तरीके से रख सकते हैं ।

सारणी संख्या 5-1

घटना	घटना की प्रायिकता	मृतक अनुपात%
1	2	3
दोनों रोगियों की मृत्यु	$\frac{1}{10000}$	100
एक की मृत्यु और एक का आरोग्य लाभ	$\frac{99}{5000}$	50
दोनों का आरोग्य-लाभ	$\frac{9801}{10000}$	0

इन तीनों घटनाओं में से केवल एक ही ऐसी है जिसमें प्रायिकता इतनी कम है कि हमें इस परिकल्पना में संदेह हो जाता है कि औषध का कुछ भी प्रभाव नहीं पड़ता । यह वह घटना है जब दोनों रोगियों की मृत्यु हो जाती है । परन्तु यदि ऐसी दुर्घटना हो जाये तो यह विश्वास हो सकता है कि औषध हानि-कारक है । दोनों रोगियों का

ठीक हो जाना ही एक ऐसी घटना है जिसमें प्रतिदर्श में मृतक अनुपात अपेक्षित अनुपात से २०% कम है तथा जो इस बात का धौनक हो सकती है कि औपध लाभदायक है। परन्तु यदि औपध का कुछ भी प्रभाव न पड़े तब भी इस घटना की प्रायिकता $\frac{1}{25} = 64\%$ इतनी अधिक है कि इससे कुछ भी निष्कप निकालना असभव है।

यह स्पष्ट है कि प्रतिदर्श में रोगियों की सख्या चाहे जितनी हो यदि सभी रोगी आरोग्य लाभ कर ले तो मृतक-अनुपात प्रतिदश में दून्य प्रतिशत होगा। औपध का कुछ भी प्रभाव नहीं होता इस परिकल्पना के आधार पर परिकल्पित इन घटना की प्रायिकता यदि इतनी अधिक है कि औपध के गुणकारी प्रभाव का विश्वास दिलाने में यह असमथ है तो कोई भी अन्य घटना जिसमें कुछ व्यक्ति मर जाते हैं और कुछ व्यक्तियों को लाभ हो जाना है यह विश्वास दिला ही नहीं सकती कि औपध से इस रोग में लाभ होता है। इसलिए इतने छोटे प्रतिदश का प्रयोग करना बेकार है।

आइए, पहले हम यह मालूम करें कि प्रतिदश में रोगियों की सख्या कम से कम कितनी होनी चाहिए कि उससे औपध के गुणकारी प्रभाव का विश्वास दिलाने की संभावना सों रहे। इसमें हमें ऐसी सख्या का पता लगाना है कि सब रोगियों के आरोग्य लाभ की प्रायिकता बहुत कम हो। इतनी कम कि लोगों को विश्वास न हो कि बिना औपध-प्रभाव के ऐसी घटना घट सकती है। नीचे सारणी में कुछ प्रतिदश सख्याएँ और तत्संबधी सभी रोगियों के आरोग्य लाभ की प्रायिकताएँ दी गयी हैं।

सारणी सख्या 5 2

प्रतिदर्श-सख्या	सभी रोगियों के आरोग्य लाभ की प्रायिकता
(1)	(2)
3	$\left(\frac{4}{5}\right)^3 = \frac{64}{125} = 0.512$
4	$\left(\frac{4}{5}\right)^4 = \frac{256}{625} = 0.4096$
5	$\left(\frac{4}{5}\right)^5 = \frac{1024}{3125} = 0.32768$
10	$\left(\frac{4}{5}\right)^{10} = 0.1074$
100	$\left(\frac{4}{5}\right)^{100} = 0.000,000,000,200$

प्रतिदर्श सख्या दस तक सभी रोगियों के आरोग्य-लाभ की प्रायिकता बिना औषध के प्रभाव के भी इतनी है कि यह औषध के लाभकारी होने में विश्वास दिलाने के लिए यथेष्ट नहीं है। शायद हमें उस समय तक विश्वास नहीं हो सकेगा जब तक इस घटना की प्रायिकता ५% से कम न हो। प्रतिदर्श सख्या सौ में इस घटना की प्रायिकता इतनी कम है—अर्थात् एक अरब में दो—कि यदि वास्तव में यह घटना घटित हो जाय तो हमें पूरा भरोसा हो जायगा कि यह औषध रोग की चिकित्सा में चमत्कारी है।

आपको याद होगा कि हमने उदाहरण सौ रोगियों के प्रतिदर्श से आरम्भ किया था जिसमें दस रोगियों की मृत्यु हुई थी। प्रश्न यह है कि यदि औषध का कुछ भी प्रभाव नहीं होता तो ऐसी घटना कहां तक संभव थी। हम दस अथवा दस से कम मृत्यु की प्रायिकता का परिकलन औषध के प्रभावहीन होने की परिकल्पना पर करना चाहेंगे। इसका कलन भी उतना ही सरल है जितना कि छोटे प्रतिदर्शों में हमने पाया था। इनके फलों को सारणी के रूप में नीचे दिया है। यदि प्रयोग के इस फल से हम यह तय करते हैं कि परिकल्पना झूठी है तो यह तय है कि यदि मृत्युसख्या इससे भी कम होती तो भी हम—शायद और भी विश्वास के साथ—परिकल्पना को झूठा समझते। हम यह जानना चाहेंगे कि यदि परिकल्पना सत्य होती तो इस प्रकार की त्रुटि की—उसको दूठ मानने की—क्या प्रायिकता है। इसके लिए हमें सारणी सख्या 53 में दी हुई प्रायिकताओं का योग करना होगा। यह योग 0.0057 है। इसके साथ ही हम

सारणी सख्या 53

घटना	घटना की प्रायिकता	मृतक-अनुपात प्रतिशत
I	2	3
100 रोगियों को आरोग्य-लाभ	$(\frac{4}{5})^{100}$	0
99 को आरोग्य-लाभ व १ की मृत्यु	$(\frac{4}{5})^{99} (\frac{1}{5}) \times 100$	1
98 को आरोग्य लाभ व २ की मृत्यु	$(\frac{4}{5})^{98} (\frac{1}{5})^2 \times \binom{100}{2}$	2
97 को आरोग्य-लाभ व ३ की मृत्यु	$(\frac{4}{5})^{97} (\frac{1}{5})^3 \times \binom{100}{3}$	3
96 को आरोग्य-लाभ व ४ की मृत्यु	$(\frac{4}{5})^{96} (\frac{1}{5})^4 \times \binom{100}{4}$	4

घटना	घटना की प्रायिकता	मृतक-अनुपात प्रतिशत
1	2	3
95 को आरोग्य-लाभ व 5 की मृत्यु	$\left(\frac{4}{5}\right)^{95} \left(\frac{1}{5}\right)^5 \binom{100}{5}$	5
94 को आरोग्य लाभ व 6 की मृत्यु	$\left(\frac{4}{5}\right)^{94} \left(\frac{1}{5}\right)^6 \binom{100}{6}$	6
93 को आरोग्य-लाभ व 7 की मृत्यु	$\left(\frac{4}{5}\right)^{93} \left(\frac{1}{5}\right)^7 \binom{100}{7}$	7
92 को आरोग्य-लाभ व 8 की मृत्यु	$\left(\frac{4}{5}\right)^{92} \left(\frac{1}{5}\right)^8 \binom{100}{8}$	8
91 को आरोग्य-लाभ व 9 की मृत्यु	$\left(\frac{4}{5}\right)^{91} \left(\frac{1}{5}\right)^9 \binom{100}{9}$	9
90 को आरोग्य-लाभ व 10 की मृत्यु	$\left(\frac{4}{5}\right)^{90} \left(\frac{1}{5}\right)^{10} \binom{100}{10}$	10

यह कह सकते हैं कि यदि हम सौ-सी रोगिया के दस सहस्र प्रतिदर्शों का अवलोकन करें तो केवल 57 में ही दस अथवा उनसे कम मृत्यु मक्ष्या होगी। इस प्रकार के प्रयोग-फल से यह धारणा बनती है कि यह औषध लाभदायक है।

सारणी 53 में दी हुई ग्यारह घटनाओं की प्रायिकताओं की गणना हमने किस प्रकार की? पहली घटना में तो यह गणना बहुत ही सरल है। सौ घटनाएँ हैं जिनमें से हर एक की प्रायिकता $\left(\frac{4}{5}\right)$ है और वे एक दूसरे से स्वतंत्र हैं। इसलिए इन सब घटनाओं के होने की प्रायिकता उनकी भिन्न भिन्न प्रायिकताओं का गुणन अर्थात् $\left(\frac{4}{5}\right)^{100}$ है।

दूसरी घटना के लिए मान लीजिए कि एक विशेष रोगी A_1 तो मर जाता है और अन्य सब रोगी आरोग्य-लाभ करते हैं। इस घटना की प्रायिकता $\left(\frac{4}{5}\right)^{99} \times \left(\frac{1}{5}\right)$ है। अब हम यदि इसी प्रकार की एक अन्य घटना की प्रायिकता का कलन करें जिसमें एक अन्य रोगी A_2 तो मर जाता है और अन्य रोगियों को आरोग्य लाभ होता है तो वह भी $\left(\frac{4}{5}\right)^{99} \times \left(\frac{1}{5}\right)$ होगी। कौन सा विशेष रोगी मरता है इस पर निर्भर कुल एक सौ घटनाएँ हैं जिनकी प्रायिकताएँ $\left(\frac{4}{5}\right)^{99} \times \left(\frac{1}{5}\right)$ हैं। इसलिए इनमें से किसी घटना के होने की—सौ में से किसी एक रोगी के मरने की—प्रायिकता $\left(\frac{4}{5}\right)^{99} \times \left(\frac{1}{5}\right) \times 100$ है।

इसी प्रकार मान लीजिए कि दो विशेष रोगी A_1 और A_2 तो मर जाते हैं तथा अन्य सब ठीक हो जाते हैं। इस घटना की प्रायिकता $(\frac{4}{5})^{99} \times (\frac{1}{5})^2$ है। हम यह भी जानते हैं कि सौ रोगियों में से दो रोगियों के $(1 \ 0 \ 0)$ कुलक (sets) बनाये जा सकते हैं। इनमें से यदि किसी विशेष कुलक के रोगी मर जाये तथा अन्य सबको आरोग्य-लाभ हो तो इसकी प्रायिकता, जैसे हम ऊपर देख चुके हैं, $(\frac{4}{5})^{99} \times (\frac{1}{5})^2$ है। इसलिए कुल प्रायिकता कि कोई भी दो रोगी मर जाये और अन्य आरोग्य-लाभ करे $(\frac{4}{5})^{99} \times (\frac{1}{5})^2 \times (1 \ 0 \ 0)$ है। इसी प्रकार के तर्क से अन्य सब प्रायिकताओं का बलन किया जा सकता है।

अध्याय ६

द्विपद वंटन (Binomial Distribution)

§ ६१ द्विपद वंटन

पिछले अध्याय के अन्त में दी हुई प्रायिकताओं के गणन का एक व्यापक सूत्र है जिसको धतुर पाठक कदाचित् अब तक मालूम भी कर चुका होगा। मान लीजिए कि एक यादृच्छिक प्रयोग (random experiment) के दो ही फल हो सकते हैं A और A' जिनमें A की प्रायिकता p है और A' की प्रायिकता $1-p=q$ है। यदि इस यादृच्छिक प्रयोग को N बार दोहराया जाये तो इस घटना की प्रायिकता कि n बार A और $N-n$ बार A' घटित हो $\binom{N}{n} p^n q^{(N-n)}$ है। प्रयोग को N बार दोहराने से जितनी बार A घटित हो वह सख्या एक यादृच्छिक चर है। इस चर का मान n होने की प्रायिकता $\binom{N}{n} p^n q^{(N-n)}$ है। यही हमारे यादृच्छिक चर का वंटन है।

यह वंटन द्विपद वंटन के नाम से विख्यात है। इसका कारण यह है कि A के घटने की भिन्न भिन्न सख्याओं की प्रायिकताएँ $(p+q)^N$ के द्विपद विस्तार से प्राप्त होती है। $(p+q)^N$ का द्विपद विस्तार निम्नलिखित है—

$$(p+q)^N = q^N + \binom{N}{1} q^{N-1} p + \binom{N}{2} q^{N-2} p^2 + \dots + \binom{N}{n} q^{N-n} p^n \\ + \dots + \binom{N}{N-2} q^2 p^{N-2} + \binom{N}{N-1} q^1 p^{N-1} + \binom{N}{N} p^N$$

इस बहुत ही महत्वपूर्ण और साधारण द्विपद वंटन के कुछ और उदाहरणों पर अब हम विचार करेंगे।

§ ६२ द्विपद वंटन के उपयोग के कुछ उदाहरण

(१) प्रायः सभी पाठक इस कहावत से परिचित होंगे कि “भूल करना मनुष्य का स्वभाव है।” कुशल से कुशल व्यक्ति भी कहीं न कहीं त्रुटि कर ही बैठते हैं। वे इसी अर्थ में कुशल माने जाते हैं कि नौसिलियों की अपेक्षा उनकी त्रुटियों की वारवारता बहुत कम होती है। एक टाइपिस्ट का विचार कीजिए—चाहे उसे टकन (type) करते हुए दस वर्ष वीत गये हों, पर यह असंभव है कि टकन करने में उसकी कभी त्रुटि नहीं होती। विशेष रूप से विचार करने के लिए मान लीजिए कि किसी एक पृष्ठ पर कम से कम त्रुटि होने की प्रायिकता पच्चीस प्रतिशत है—अर्थात् यदि हम टकन किये हुए अनेक पृष्ठों की परीक्षा करें तो उनमें लगभग एक चौथाई में एक या अधिक त्रुटियाँ होंगी। अब यदि यह दशा एक अनुभवशील टाइपिस्ट की है तो नये व्यक्ति से इससे कम त्रुटियाँ करने की आशा करना व्यर्थ है। यदि यह अनुभवशील टाइपिस्ट नौकरी छोड़ कर जा रहा हो और मैनेजर को नये आदमी की नियुक्ति करनी हो तो वह यह जानना चाहेगा कि प्रार्थी की योग्यता लगभग उस व्यक्ति के बराबर है या नहीं जो नौकरी छोड़ रहा है। यदि वह अधिक योग्य हो या लगभग बराबर योग्यता रखता हो तो नौकरी देने में कुछ आपत्ति नहीं होनी चाहिए। परन्तु यदि उसकी योग्यता बहुत कम है तो अधिक त्रुटियाँ होने के कारण काम का समय अधिक नष्ट होगा। यह जानने के लिए कि प्रार्थी की योग्यता कितनी है—एक ही तरीका है—वह यह कि उससे टकन करवा कर परीक्षा ली जाये। मान लीजिए कि परीक्षा के लिए टाइपिस्ट को चालीस पृष्ठ टकन के लिए दिये जाते हैं। परिकल्पना यह है कि प्रार्थी औसतन उस व्यक्ति से अधिक त्रुटि नहीं करता जो नौकरी छोड़कर जा रहा है। इस आधार पर हमें प्रयोग में प्रेषित त्रुटियों की संख्या के बराबर और उनसे अधिक त्रुटियों की प्रायिकता की गणना करना है।

यदि इन प्रयोग में दस से कम पृष्ठों में ही त्रुटि पायी जाती है तो स्पष्टतः टकन उस औसत मान से अपेक्षाकृत अधिक अच्छा है जिसकी हम आशा करते थे। तब तो हमें प्रायिकता का कलन करने की कोई आवश्यकता नहीं है। यह आवश्यकता उसी समय पड़ेगी जब परिणाम औसत में खराब हो। आइये, हम देखें कि एक ऐसे प्रार्थी के बारे में मैनेजर का क्या निर्णय होना चाहिए जो इस प्रयोग में 13 पृष्ठों की त्रुटियों के कारण बिगाड़ देता है।

यदि आप मैनेजर हैं तो आप यह तो देखेंगे ही कि परिणाम आशा से खराब है, परन्तु आप यह भी जानते हैं कि ऐसा केवल संयोग से होना भी संभव है, यदि २५%

पर त्रुटियाँ की परिकल्पना पर आधारित प्रायिकता तेरह पृष्ठों पर भूलों के लिए काफी है तो न्यायसोल होने के नाते आप प्रार्थी को असफल घोषित करना ठीक नहीं समझेंगे। शायद आप उसकी परीक्षा को और बढ़ा दें तब उसे कुछ अधिक पृष्ठ टाइप करने को दें जिससे आप अधिक निःसंकोच होकर निर्णय कर सकें।

आइये, अब चालीस पृष्ठों में से तेरह अथवा तेरह से अधिक पर त्रुटियाँ होने की प्रायिकता की गणना की जाये। इसमें हमें बट्टाडस भिन्न भिन्न प्रायिकताओं की गणना करके उनका योग करना होगा। परन्तु हम इसी को एक दूसरे ढंग से भी हल कर सकते हैं जिसमें मेहनत कम हो।

P (चालीस में से तेरह अथवा उससे भी अधिक पृष्ठों पर त्रुटियाँ होना)

= 1 - P (चालीस में से बारह अथवा उससे भी कम पृष्ठों पर त्रुटियाँ होना)

अब बारह अथवा उससे भी कम पृष्ठों पर त्रुटियाँ होने की प्रायिकता का कलन करने के लिए केवल तेरह आरम्भ घटनाओं की प्रायिकताओं का कलन करने और उनका योग करने की आवश्यकता है। यह गणना अगले पृष्ठ की सारणी में दी हुई है।

इसलिए बारह अथवा इससे कम त्रुटियाँ के होने की कुल प्रायिकता

$$= \frac{3^{28}}{4^{40}} \left\{ \binom{40}{12} + 3 \binom{40}{11} + 3^2 \binom{40}{10} + \dots + 3^{12} \right\}$$

$$= 0.8208658$$

∴ तेरह अथवा तेरह से अधिक त्रुटियों की प्रायिकता

$$= 1 - 0.8208658$$

$$= 0.1791342$$

इस प्रकार हम देखते हैं कि "किसी पृष्ठ पर त्रुटि होने की प्रायिकता पच्चीस प्रतिशत अर्थात् 0.25 है" ऐसी परिकल्पना के आधार पर प्रयोग के फल की प्रायिकता इतनी कम नहीं है कि हम परिकल्पना को त्यागने के लिए बाध्य हो जायें और हमारा यह विश्वास हो जाये कि प्रार्थी के लिए किसी पृष्ठ पर त्रुटि होने की प्रायिकता अवश्य पच्चीस प्रतिशत से अधिक होगी। इस दशा में मंजूर उसे नियुक्त करना अनुचित नहीं समझेंगे।

(२) द्विपद बटन का उपयोग केवल औपधिषा के गुण की परीक्षा अथवा नौकरी के लिए उपयुक्त व्यक्तियों के चुनाव तक ही सीमित नहीं है। शायद इसका सबसे अधिक उपयोग व्यापार में माल के स्वीकार अथवा अस्वीकार करने में होता है। पुस्तक के आरम्भ में ही हम यह देख चुके हैं कि साधारणतया मनुष्य प्रतिद्वंद्व के आधार पर ही

सारणी सख्या 61

घटना	प्रायिकता
(1)	(2)
किसी पृष्ठ पर त्रुटि नहीं है	$\left(\frac{3}{4}\right)^{40}$
केवल एक पृष्ठ पर त्रुटि है	$\binom{40}{1} \left(\frac{3}{4}\right)^{39} \left(\frac{1}{4}\right)$
केवल दो पृष्ठों पर त्रुटि है	$\binom{40}{2} \left(\frac{3}{4}\right)^{38} \left(\frac{1}{4}\right)^2$
केवल तीन पृष्ठों पर त्रुटि है	$\binom{40}{3} \left(\frac{3}{4}\right)^{37} \left(\frac{1}{4}\right)^3$
केवल चार पृष्ठों पर त्रुटि है	$\binom{40}{4} \left(\frac{3}{4}\right)^{36} \left(\frac{1}{4}\right)^4$
केवल पांच पृष्ठों पर त्रुटि है	$\binom{40}{5} \left(\frac{3}{4}\right)^{35} \left(\frac{1}{4}\right)^5$
केवल छ पृष्ठों पर त्रुटि है	$\binom{40}{6} \left(\frac{3}{4}\right)^{34} \left(\frac{1}{4}\right)^6$
केवल सात पृष्ठों पर त्रुटि है	$\binom{40}{7} \left(\frac{3}{4}\right)^{33} \left(\frac{1}{4}\right)^7$
केवल आठ पृष्ठों पर त्रुटि है	$\binom{40}{8} \left(\frac{3}{4}\right)^{32} \left(\frac{1}{4}\right)^8$
केवल नौ पृष्ठों पर त्रुटि है	$\binom{40}{9} \left(\frac{3}{4}\right)^{31} \left(\frac{1}{4}\right)^9$
केवल दस पृष्ठों पर त्रुटि है	$\binom{40}{10} \left(\frac{3}{4}\right)^{30} \left(\frac{1}{4}\right)^{10}$
केवल ग्यारह पृष्ठों पर त्रुटि है	$\binom{40}{11} \left(\frac{3}{4}\right)^{29} \left(\frac{1}{4}\right)^{11}$
केवल बारह पृष्ठों पर त्रुटि है	$\binom{40}{12} \left(\frac{3}{4}\right)^{28} \left(\frac{1}{4}\right)^{12}$

केय विक्रय करते हैं। लेकिन यह बहुत कुछ अनुमान पर आधारित होता है। एक बड़ा व्यापारी जो कारखानों से बड़े पैमाने पर माल खरीदता है इस अनुमान को बेजा निक रीति से लगाना चाहेगा कि जिससे उसे अधिक से अधिक लाभ हो। एक बार में उसे जो माल मिलता है उसे डेरी (lot) कहते हैं। यद्यपि कारखाना में ये वस्तुएँ

मशीनों से बनती है, तथापि एक ही डेरी की भिन्न-भिन्न वस्तुओं में भी अंतर पाया जाता है। कारखाने की भिन्न-भिन्न मशीनों में अंतर, मशीनों के समजन (adjustment) से पड़ने वाला अन्तर, कच्चे माल में अन्तर आदि कुछ ऐसे कारण हैं जिनसे अंत में कारखाने से निकली वस्तुओं में अन्तर पड़ जाता है। बलों के उपयोग करनेवाले मजदूरों की चतुरता पर भी यह बहुत कुछ निर्भर करता है।

यदि यह अंतर साधारण-सा हो तो व्यापारी इसकी उपेक्षा कर देगा क्योंकि ग्राहक या तो इस अन्तर को पहचान ही नहीं पायेंगे या उसको कोई विशेष महत्त्व नहीं देंगे। परन्तु यह संभव है कि यह अन्तर इतना स्पष्ट हो उठे कि ग्राहक वस्तु खरीदना अस्वीकार कर दे। ऐसी वस्तुओं को दोषपूर्ण मानना होगा। कारखाने के लिए दो रास्ते हैं—एक तो यह कि वह डेरी में से प्रत्येक वस्तु का निरीक्षण करके उनमें से दोषयुक्त वस्तुओं को निकाल दे। इस प्रकार वे माल के दश प्रतिशत अच्छे होने की प्रतिश्रुति (guarantee) दे सकते हैं। लेकिन इस तरीके में दो कठिनाइयाँ हैं। पहली तो यह कि हर एक वस्तु के निरीक्षण से भी विलकुल निश्चयपूर्वक नहीं कह सकते कि हर वस्तु ठीक ही है। इस कथन पर पहले तो आपको आश्चर्य होगा। परन्तु निरीक्षण तो मनुष्य द्वारा ही होता है और मनुष्य से गलती होना स्वाभाविक ही है। यदि एक मनुष्य सैकड़ों वस्तुओं का निरीक्षण कर चुका है और वह सब दोषरहित हैं तो यह स्वाभाविक है कि शेष वस्तुओं का निरीक्षण उतनी बारीकी से नहीं होगा। यह भी संभव है कि वह कई वस्तुओं को बिना यथेष्ट परीक्षण के ही स्वीकार कर ले। दूसरी कठिनाई यह है कि इस निरीक्षण से व्यय बढ़ जाता है।

मान लीजिए कि एक डेरी में दस हजार वस्तुएँ हैं जिनकी कुल कीमत एक लाख रुपये है और इनमें से एक प्रतिशत दोषयुक्त है। इसका यह अर्थ हुआ कि व्यापारी एक हजार रुपये की वस्तुएँ नहीं बेच पायेगा। और यदि उसने बेच भी दी तो संभवतः उसे उनको वापिस लेकर दोषरहित वस्तुओं से बदलना पड़े। यदि इस हानि से बचने के लिए कारखाना या व्यापारी पूर्ण निरीक्षण का प्रयोग करे जिसमें उसको एक हजार रुपये से अधिक का खर्च पड़ जाये तो इस निरीक्षण का कोई विशेष लाभ नहीं है। कुल व्यय का हिसाब लगाकर व्यापारी डेरी में कुछ प्रतिशत दोषयुक्त वस्तुओं को सहन करना स्वीकार कर लेगा।

दूसरा रास्ता उसके लिए प्रतिदर्श पर निर्भर करता है। प्रतिदर्श कितना बड़ा होना चाहिए, यह इस पर निर्भर करता है कि व्यापारी को कितनी प्रतिशत दोषयुक्त वस्तुएँ स्वीकार करना सहन है। यदि हम त्रुटि की इस चरम प्रतिशतता को P से

सूचित करें तो हमारी सांख्यिकीय समस्या केवल इस परिकल्पना की जाँच करना है कि डेरी में P प्रतिशत वस्तुएँ या P प्रतिशत से कम वस्तुएँ दोषयुक्त हैं। यदि प्रतिशत में दोषयुक्त वस्तुओं का अनुपात P से अधिक P' हो और उपर्युक्त परिकल्पना के आधार पर परिकलित इस घटना की प्रायिकता बहुत कम हो कि प्रतिदर्श में P' प्रतिशत अथवा उससे अधिक वस्तुएँ दोषयुक्त हैं तो हम यह समझेंगे कि उम परिकल्पना को इस प्रयोग के आधार पर अस्वीकृत कर देना चाहिए और यह मानना चाहिए कि वास्तव में डेरी में दोषयुक्त वस्तुओं का अनुपात P से अधिक है। इस दशा में डेरी को अस्वीकार करना ही युक्तिसंगत है। क्योंकि P प्रतिशत ही वह पराकाष्ठा है जहाँ तक वह डेरी का दूषित होना सहन कर सकता है।

स्पष्टतया इस उदाहरण में तथा पिछले उदाहरण में, जिसमें प्राथियों के चुनाव की समस्या थी, बहुत अधिक समानता है। वास्तव में वैज्ञानिक अनुसंधानों और प्रतिदिन के जीवन में, क्रय-विक्रय में, योग्य व्यक्तियों के निर्वाचन में, तथा नये-नये साधना की कार्य-साधकता की परीक्षा में सैकड़ा ऐसे उदाहरण हमारे सामने आते हैं जिनमें हम यह जानना चाहते हैं कि कोई विशेष प्रयोगलब्ध अनुपात किसी दी हुई सख्या से बड़ा है अथवा नहीं। इन सब स्थितियों में प्रायिकताओं की गणना द्विपद वटन की महायता से की जाती है।

§ ६.३ द्विपद वटन के कुछ गुण

पाठकों को इस महत्त्वपूर्ण वटन के बारे में अधिक जानकारी करने की उत्सुकता अवश्य होगी। इसके कुछ गुणों का वर्णन नीचे दिया गया है —

(१) यह वटन असतत है। यदि प्रतिदर्श-संख्या N है तो द्विपद चर केवल $(N+1)$ भिन्न भिन्न मान धारण कर सकता है जो $0, 1, 2, 3, 4, \dots, n, N$ हैं।

(२) इस चर का मान n होने की प्रायिकता $\binom{N}{n} p^n (1-p)^{N-n}$ है। p शून्य व एक के बीच की कोई सख्या है। इस प्रकार N और p दो ऐसे मान हैं जिनसे विशेष द्विपद वटन निर्दिष्ट हो जाता है।

(३) इसका वटन-फलन (distribution function) याने n अथवा n से कम मान धारण करने की प्रायिकता $F(n) = \sum_{x=0}^n \binom{N}{x} p^x (1-p)^{N-x}$ है।

(४) परिभाषा के अनुसार इस वटन का माध्य अथवा प्रत्याशित मान

$$\begin{aligned}
 \mu(n) = E(n) &= \sum_{n=0}^N n_x \binom{N}{n} p^n q^{N-n} \\
 &= \sum_{n=0}^N n_x \frac{N!}{n!(N-n)!} p^n q^{N-n} \\
 &= Np \sum_{n=1}^N \frac{(N-1)!}{(n-1)!(N-n)!} p^{n-1} q^{N-n} \\
 &= Np (p+q)^{N-1} \\
 &= Np
 \end{aligned} \tag{6.1}$$

द्वारा कि $p+q=1$

(5) इसी प्रकार इस वटन का प्रसरण

$$\begin{aligned}
 \sigma^2(n) &= E[n - E(n)]^2 \\
 &= E[n^2 - 2nE(n) + E^2(n)] \\
 &= E(n^2) - E^2(n)
 \end{aligned}$$

$$\begin{aligned}
 E(n^2) &= \sum_{n=0}^N n^2 \binom{N}{n} p^n q^{N-n} \\
 &= \sum_{n=0}^N \{n(n-1) + n\} \frac{N!}{n!(N-n)!} p^n q^{N-n} \\
 &= N(N-1)p^2 \sum_{n=2}^N \binom{N-2}{n-2} p^{n-2} q^{N-n} \\
 &\quad + Np \sum_{n=1}^N \binom{N-1}{n-1} p^{n-1} q^{N-n} \\
 &= N(N-1)p(p+q)^{N-2} + Np(p+q)^{N-1} \\
 &\quad (p+q) + Np
 \end{aligned}$$

$$E(n) = Np$$

$$\begin{aligned}
 \text{इसलिए } \sigma^2(n) &= N(N-1)p^2 + Np - N^2p \\
 &= Np - Np \\
 &= Np(1-p) \\
 &= Npq
 \end{aligned}$$

हम इस वंटन के सभी घूर्णों का उपर्युक्त रीति से परिकलन कर सकते हैं। यह रीति अब तक पाठको को स्पष्ट हो गयी होगी। इसलिए और अधिक घूर्णों की गणना करना यहाँ आवश्यक नहीं है। प्रथम दो पूर्ण माध्य व विचरण जिनका परिकलन ऊपर किया गया है अधिक महत्त्व रखते हैं, जैसा कि आगे हमें विदित होगा। इसके अतिरिक्त इस वंटन के अन्य गुण जैसे माध्यिका (median), चतुर्थक (quartiles) दशमक (deciles) या शततमक (percentiles) भी वंटन की सभी घटनाओं के ज्ञात होने के कारण परिकलित किये जा सकते हैं, किन्तु क्योंकि यह एक असतत वंटन है इसलिए परिभाषा के अनुसार यह बहुत संभव है कि कई गुण वंटन में विद्यमान न हों। मान लीजिए $N=2$ और $p=\frac{1}{2}$ । इस स्थिति में μ केवल तीन मान धारण करता है 0, 1 और 2। इनको धारण करने की प्रायिकताएँ क्रमशः $\frac{1}{2}$, $\frac{1}{2}$ और $\frac{1}{2}$ हैं। इस वंटन में कोई भी ऐसी सख्या नहीं है जिसके बराबर या उससे कम मान धारण करने की प्रायिकता $\frac{1}{2}$ हो। इस प्रकार परिभाषा के अनुसार इस वंटन में कोई माध्यिका नहीं है। हम चाहें तो इसको 0 और 1 के बीच की कोई सख्या मान सकते हैं क्योंकि 0 मान धारण करने की प्रायिकता $\frac{1}{2} < \frac{1}{2}$ और 1 या 2 धारण करने की प्रायिकता $\frac{1}{2} > \frac{1}{2}$ है।

परन्तु इसी तर्क से यह माध्यिका 1 और 2 के बीच की कोई सख्या भी हो सकती है। इस प्रकार किसी सखेच्छ नियम द्वारा यद्यपि माध्यिका की परिभाषा दी जा सकती है, परन्तु उसका कोई प्रयोग महत्त्व नहीं होगा। जिस प्रकार इस द्विपद वंटन में माध्यिका का अस्तित्व नहीं है उसी प्रकार इनमें और अन्य कई द्विपद वंटनों में दशमक, शततमक आदि का अस्तित्व नहीं होता। इसी कारण ये गुण इतने अधिक महत्त्वपूर्ण नहीं समझे गये हैं तथा इनके परिकलन के लिए व्यर्थ खेपटा यहाँ नहीं की गयी है।

§ ६४ द्विपद-वंटन के लिए सारणी

इस वंटन का बहुत ही व्यापक प्रयोग होने के कारण संभव है कि एक ही N और p के मानवाले वंटनों का अनेक वैज्ञानिक भिन्न-भिन्न स्थितियों में तथा भिन्न-भिन्न देशों में उपयोग करते होंगे। इन सबको बार-बार एक ही प्रकार का परिकलन यदि केवल यह ज्ञानने के लिए करना पड़े कि प्रयोग के फल को देखते हुए परिष्कृत्यना को स्वीकार करना चाहिए अथवा नहीं, तब यह मानसिक शक्तियों का अपभ्रंश होगा। क्या यह नहीं हो सकता कि जिस किसीने एक बार एक विशेष वंटन के लिए परिकलन किया हो वह उसको अपनी और दूसरों की बृथा मेहनत बचाने के लिए अभिलेख-बद्ध

वर ले और प्रकाशित करा दे ? इसी विचार से सांख्यिकी ने इस बटन की सारणी तैयार की है जिसमें

$$F(n) = \sum_{x=0}^n \binom{N}{x} p^x q^{n-x} \quad (63)$$

के मान N के एक से लेकर पचास तक के, n के शून्य से लेकर N तक के और P के 0.0, 0.0201, 0.03, . . ., 0.98, 0.99, 1.00 मानों के लिए दे रखे हैं। दो उदाहरण नीचे दिये हुए हैं।

सारणी सख्या 62

दो द्विपद-बटनों के संचित प्रायिकता फलन

$$N=25$$

$$p=0.50$$

r	$F(r)$
(1)	(2)
1	0.0000008
2	0.0000097
3	0.0000783
4	0.0004553
5	0.0020387
6	0.0073166
7	0.0216426
8	0.0538761
9	0.1147615
10	0.2121781
11	0.3450190
12	0.5000000

r	$F(r)$
(1)	(2)
13	0.6549810
14	0.7878219
15	0.8852385
16	0.9461239
17	0.9783574
18	0.9926834
19	0.9979613
20	0.9995447
21	0.9999217
22	0.9999903
23	0.9999992
24	1.0000000

$N=40$

r	$F(r)$
(1)	(2)
14	0 0000001
15	0 0000006
16	0 0000028
17	0 0000123
18	0 0000486
19	0 0001749
20	0 0005724
21	0 0017084
22	■ 0046515
23	0 0115614
24	■ 0262449
25	■ 0544372

 $p=0.25$

r	$F(r)$
(1)	(2)
26	0 1032317
27	0 1791342
28	0 2848556
29	0 4160959
30	0 5604603
31	0 7001677
32	0 8180458
33	0 9037754
34	0 9567260
35	■ 9839578
36	■ 9953043
37	0 9989843

विस्तृत मासणी के लिए देखिए—“Tables of the Incomplete Beta-Function” by Karl Pearson

§ ६.५ एक मनोवैज्ञानिक सिद्धान्त की जाँच में द्विपद वटन का उपयोग

हम इस अध्याय को एक मनोवैज्ञानिक प्रयोग के विवरण में समाप्त करेंगे जिसमें इस वटन का प्रयोग होता है :

एक ही काम करने के कई ढंग हो सकते हैं । संभव है कि एक ही मनुष्य को यह सब ढंग ज्ञात हो । यदि उसके पास सोचने का काफी समय है और मस्तिष्क-शक्ति स्वस्थ है तो वह अवश्य ही इनमें से सबसे सरल पद्धति को अपनायेगा । यह एक मनोवैज्ञानिक सिद्धान्त है कि यदि मनुष्य थका हुआ हो अथवा उसके पास सोचने का अधिक अवकाश न हो तो वह कार्य करने की उस पद्धति को अपनायेगा जिसको उसने सबसे प्रथम सीखा था । यह केवल एक सांख्यिकीय कथन है । इसका यह दावा नहीं है कि प्रत्येक मनुष्य प्रत्येक बार जब ऐसी स्थिति होगी तो इस ही प्रकार आचरण करेगा । यह केवल यह बताता है कि अधिकतम मनुष्य उसी तरीके को अपनायेंगे जिसे उन्होंने पहिले सीखा हो ।

समस्या है इस प्रयोग द्वारा सिद्धान्त की परीक्षा करने की । कालेज के अठारह विद्यार्थियों को गुणा करने के दो तरीके सिखाये गये । इनमें से नौ को पहला तरीका प्रथम और दोप नौ को दूसरा तरीका प्रथम सिखाया गया । एक दिन छ घंटे के कठिन मानसिक परिश्रम के पश्चान् उनको गुणा करने के लिए कुछ प्रश्न दिये गये । सिद्धान्त के अनुसार यह आशा की जाती थी कि थकान के कारण ये विद्यार्थी उस तरीके का उपयोग करेंगे जिसको उन्होंने पहले सीखा था । प्रत्येक विद्यार्थी की दो श्रेणियों में से एक में रज दिया गया । एक श्रेणी तो उन विद्यार्थियों की थी जिन्होंने प्रथम सीखे हुए तरीके का उपयोग किया, दूसरे वे जिन्होंने बाद में सीखे हुए तरीके का उपयोग किया ।

यह परिकल्पना जिसकी हम परीक्षा करेंगे यह है कि पहले और बाद में सीखे हुए तरीका को इस स्थिति में अपनाने की प्रायिकताएँ बराबर हैं अर्थात् दोनों प्रायिकताएँ $\frac{1}{2}$ हैं । यदि प्रतिदर्श में इन दो श्रेणियों के अनुपात की संख्या बराबर न भी हो तो उनमें अंतर इतना ही होना चाहिए कि यह समझा जा सके कि यह केवल संयोग का फल है । प्रेक्षित अंतर अथवा उससे भी अधिक अंतर की प्रायिकता इतनी कम नहीं होनी चाहिए कि हम अपनी परिकल्पना से सदेह होने लगे । यदि यह अंतर अधिक है और हम परिकल्पना को अस्वीकार करते हैं तो हम यह भी कह सकते हैं कि इस सिद्धान्त की पुष्टि होती है कि थकान की दशा में प्रथम सीखे हुए तरीके के अपनाये जाने की अधिक प्रायिकता है ।

प्रयोग में देखा गया कि केवल दो विद्यार्थियों को छोड़कर बाकी सबने पहले सीखे हुए तरीके का उपयोग किया। ये आँकड़े नीचे सारणी में दिये हुए हैं।

सारणी संख्या 63

	पद्धति जो अपनायी गयी		कुल
	पहिले सीखी हुई	बाद में सीखी हुई	
(1)	(2)	(3)	(4)
वारवारता	16	2	18

इस प्रेक्षित अंतर और इससे अधिक अन्तर की प्रायिकता के कलन नीचे दिये हुए हैं।

सारणी संख्या 64

घटना	प्रायिकता
(1)	(2)
16 पहली श्रेणी और 2 दूसरी श्रेणी में	$\binom{18}{2} \left(\frac{1}{2}\right)^{18}$
17 पहली श्रेणी और 1 दूसरी श्रेणी में	$\binom{18}{1} \left(\frac{1}{2}\right)^{18}$
18 पहली श्रेणी और दूसरी श्रेणी में कोई नहीं	$\left(\frac{1}{2}\right)^{18}$

इसलिए 16 या उससे अधिक विद्यार्थियों के प्रथम श्रेणी में होने की प्रायिकता

$$\begin{aligned}
 P &= \left(\frac{1}{2}\right)^{18} \left\{ \frac{18 \times 17}{1 \times 2} + 18 + 1 \right\} \\
 &= \frac{182}{2^{18}} \\
 &= \frac{91}{131072} \\
 &< 0.001
 \end{aligned}$$

क्योंकि यह प्रायिकता एक हजार में से एक से भी कम है, हमें उस आधार पर संदेह होना स्वाभाविक ही है जिससे इस प्रायिकता का कलन किया गया है और इस कारण हम परिकल्पना को अस्वीकार करते हैं। इसका विकल्प यह है कि प्रयोग से सिद्धान्त की पुष्टि होती है।

इस अध्याय में हमने केवल द्विपद घटन के उपयोग पर विचार किया है जिससे कुछ घटनाओं की प्रायिकताओं का परिकलन किया जा सकता है। इसमें परिकल्पना की जाँच केवल प्रासंगिक थी। अगले दो अध्यायों में हम कुछ अन्य घटनों का अध्ययन करेंगे और उदाहरणों द्वारा उनके उपयोग को समझेंगे। इसके पश्चात् ही हम परिकल्पना की जाँच के सिद्धान्त (theory of testing of hypothesis), प्रतिदर्श-समस्या का निश्चित करना इत्यादि अन्य संबंधित समस्याओं पर विस्तारपूर्वक विचार करेंगे।

अध्याय ७

प्वासों-वंटन (Poisson's distribution)

§ ७.१ कुछ परिस्थितियाँ, जिनमें प्वासो-वंटन का उपयोग होता है

पिछले अध्याय में जब हम द्विपद वंटन के उपयोग पर विचार कर रहे थे, तब हमने एक निर्दिष्ट प्रतिदर्श समस्या ली थी और हमें ज्ञात था कि उसमें एक विशेष घटना कितनी बार होती है, और यह भी ज्ञात था कि वह घटना कितनी बार नहीं होती। उदाहरण के लिए टाइपिस्ट की परीक्षा के लिए हमने देखा था कि चालीस पृष्ठों में से तेरह पृष्ठों पर त्रुटियाँ थीं व सत्ताईस पृष्ठों पर कोई गलती न थी। किसी औपध के लाभदायक गुण की परीक्षा के लिए हमने यह गणना की थी कि कितने रोगी आरोग्य लाभ कर लेते हैं और कितने ठीक नहीं होते।

परन्तु ऐसे भी कई प्रयोग हैं जहाँ यद्यपि हम यह तो गिन सकते हैं कि घटना कितनी बार होती है, परन्तु उसके न होने की संख्या इतनी अधिक होती है कि उसके गिनने की परेशानी से हम बचना चाहेंगे। टाइपिस्ट की परीक्षा को ही एक दूसरे दृष्टिकोण से देखा जा सकता है। कल्पना कीजिए कि टंकित पृष्ठ पर लगभग साढ़े-चार सौ शब्द हैं, जिनमें लगभग अठारह सौ अक्षर हैं और औसत से एक पृष्ठ पर केवल 1.5 त्रुटियाँ हो सकती हैं। इसका यह अर्थ है कि एक अक्षर के गलत टंकित होने की प्रायिकता प्राय $\frac{1.5}{1800}$ है। इस दृष्टा में गलतियों की भिन्न-भिन्न संख्याओं की प्रायिकता के

परिकलन में द्विपद वंटन के उपयोग में दो कठिनाइयाँ हैं। एक तो यह कि इतनी कम प्रायिकता और इतनी अधिक प्रतिदर्श संख्या के लिए पहले से परिकल्पित द्विपद वंटन की सारणी प्रस्तुत नहीं है। इस कारण इस प्रकार के हर प्रयोग में नये सिरे से परिकल्पन आवश्यक होगा। दूसरी कठिनाई, जो मैथान्तिक रूप से अधिक महत्त्वपूर्ण है, यह है कि प्रत्येक पृष्ठ पर अक्षरों की संख्या ठीक अठारह सौ तो नहीं है। किसी पृष्ठ पर वह केवल 1760 हो सकती है, जब कि अन्य किसी पृष्ठ पर 1840 तक पहुँच सकती

है। हमारा प्रतिदर्श एक पृष्ठ है, न कि अठारह मी अक्षरो का एक समूह। द्विपद वटन इस बात पर आधारित है कि प्रतिदर्श-संख्या निश्चित हो।

इसी प्रकार एक व्यापारी दिन में 25 बार अपने टेलीफोन का प्रयोग करता है। इन प्रयोगों में, जो सक्षिप्त समाचार भेजने के लिए किये जाते हैं, समय बहुत कम या लगभग नहीं के बराबर लगता है। इस घटना की प्रायिकता कि किसी एक विशेष क्षण पर व्यापारी अपने फोन का प्रयोग कर रहा होगा, लगभग शून्य है। फिर भी दिन भर में इतने अधिक क्षण होते हैं कि पूरे दिन में हम औसतन 25 समाचार भेजे जाने की ही आशा करते हैं।

जब एक डाक्टर कीटाणुनाश या वैकटीरिया की बीजदमी का पता लगाने के लिए किसी रोगी के रक्त की परीक्षा करता है, तो उसकी विधि सद्योप में निम्नलिखित है। रक्त की बंद को एक पतली काँच की पट्टी पर फैला लिया जाता है। यह पट्टी अनेक छोटे वर्गों में विभाजित होती है। व्याधिविज्ञ इनमें से कुछ वर्गों में कीटाणुओं की गणना करता है। कुछ थोड़े से वर्गों में कीटाणुओं की गणना की जा सकती है, परन्तु बड़ाचित् कुल वर्गों के कीटाणुओं की गिनना कठिन है। इसी प्रकार कारखाने की तैयार वस्तुओं में त्रुटियों की गणना की जा सकती है पर अ-त्रुटियों की नहीं।

इन सभी अवस्थाओं में, यादृच्छिक प्रयोग की प्रतिदर्श संख्या या तो बहुत बड़ी और ज्ञात होती है अथवा इतनी बड़ी होती है कि उसका जानना ही कठिन है। साथ ही साथ प्राथमिक घटनाओं की प्रायिकता बहुत ही छोटी, शून्यप्राय ही होती है, लेकिन प्रतिदर्श-संख्या के बड़े होने के कारण प्रतिदर्श में उस घटना के होने की प्रायिकता इतनी छोटी और शून्यप्राय नहीं होती। अतः हम द्विपद वटन का प्रयोग छोड़कर एक दूसरे प्रकार का वटन अपनाते हैं। यह वटन भी द्विपद वटन से ही व्युत्पन्न है।

§ ७२ द्विपदवटन का सीमान्त रूप

हम इस प्रकार के N और p के अनेक मानों की कल्पना कर सकते हैं, जिनका गुणनफल 15 है। जैसे $N=3$, $p=\frac{1}{2}$, $N=6$, $p=\frac{1}{3}$; $N=9$, $p=\frac{1}{3}$, $N=1500$, $p=1.000$ आदि।

जैसे जैसे N का मान बढ़ता जाता है, p का मान शून्य की ओर अप्रसर होता जाता है। ये सभी मान-युग्म एक एक द्विपद की परिभाषा करते हैं, जिनमें सबके प्राचलों का गुणनफल 15 है। द्विपद चर केवल पूर्णसंख्यक मान ही धारण कर सकते हैं। किसी पूर्ण संख्या को लीजिए तो इनमें से हर एक वटन के लिए हम इस चर के इस पूर्ण

संख्या से कम अथवा बराबर मान धारण करने की प्रायिकता वा कलन कर सकते हैं। जैसे-जैसे N का मान बढ़ता जाता है, यह प्रायिकता एक निश्चित सीमान्त संख्या की ओर अग्रसर होती जाती है। हम एक ऐसे बटन की कल्पना कर सकते हैं, जिसके लिए चर के उस विशेष पूर्ण-संख्या से कम या बराबर मान धारण करने की प्रायिकता यही सीमान्त संख्या है। यह बात केवल एक विशेष पूर्ण-संख्या के लिए ही नहीं बल्कि प्रत्येक पूर्णसंख्या के लिए सत्य है। आइए, हम देखें कि इस सीमान्त बटन की परिभाषा क्या है। अर्थात् इस बटन में चर के लिए निम्नी विशेष मान r को प्राप्त करने की प्रायिकता क्या है। हम इस बटन के साधारण रूप का परिचय प्राप्त करना चाहेंगे, न कि केवल ऐसे द्विपद बटनों के सीमान्त रूप का, जिनके प्राचल N और p का गुणफल 1 हो।

यदि हम इन द्विपद बटन के माध्य को λ से सूचित करें तो प्राथमिक घटना की प्रायिकता p को $\frac{\lambda}{N}$ के बराबर रख सकते हैं। यह इसलिए कि द्विपद बटन में माध्य का मान Np होता है जैसा हम पिछले अध्याय में सिद्ध कर चुके हैं।

$$\text{अतः} \quad Np = \lambda$$

$$\therefore p = \frac{\lambda}{N}$$

इस प्रकार λ तो अचर है और सीमान्त विधि में केवल N का मान उत्तरोत्तर बढ़ता जाता है। आइये हम देखें कि उपर्युक्त बटन में चर का मान r होने की प्रायिकता क्या है।

$$P(r) = \binom{N}{r} p^r (1-p)^{N-r}$$

$$= \frac{N(N-1)(N-2)\dots(N-r+1)}{r!} \left(\frac{\lambda}{N}\right)^r \left(1 - \frac{\lambda}{N}\right)^{N-r}$$

$$= \frac{\lambda^r}{r!} \left(1 - \frac{\lambda}{N}\right)^N \times \frac{\left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right)\dots\left(1 - \frac{r-1}{N}\right)}{\left(1 - \frac{\lambda}{N}\right)^r}$$

अब यदि r के किसी निश्चित मान के लिए N का मान बढ़ता जाता है तो

$$\left(\frac{1}{N}\right), \left(1 - \frac{2}{N}\right), \left(1 - \frac{r-1}{N}\right) \text{ और } \left(1 - \frac{\lambda}{N}\right)^r$$

ये सभी संख्याएँ 1 के अधिकाधिक निवृत्त जाती जाती हैं। और $\left(1 - \frac{\lambda}{N}\right)^N$

अप्रसर होता है $e^{-\lambda}$ की ओर जहाँ

$$\begin{aligned} e &= 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \dots \\ &= 1 + \sum_{r=1}^{\infty} \frac{1}{r!} \end{aligned}$$

और z का एक विशेष गुण यह होता है कि किसी भी संख्या के लिए

$$\begin{aligned} z &= 1 + \frac{z}{1!} + \frac{z^2}{2!} + \frac{z^3}{3!} + \dots \\ &= 1 + \sum_{r=1}^{\infty} \frac{z^r}{r!} \end{aligned}$$

इस प्रकार प्रायिकता $P(r) = \frac{\lambda^r}{r!} e^{-\lambda}$ । वह घटन जिसमें चर केवल पूर्ण संख्याओं के ही बराबर हो सकता हो और प्रत्येक पूर्ण संख्या के बराबर हो सकता हो और जिसमें चर का मान किसी पूर्ण संख्या r के बराबर होने की प्रायिकता

$$P(r) = \frac{\lambda^r}{r!} e^{-\lambda} \quad (7.1)$$

हो वह प्वासा वटन के नाम से विख्यात है। पाठकों को शायद यह भ्रम हो कि इस प्रकार का वटन हो भी सकता है अथवा नहीं, इसकी परीक्षा हर एक पूर्ण संख्या से सगत प्रायिकताओं का योग करके हो सकती है। यदि यह योग 1 हो तो हम कह सकते हैं कि इस प्रकार का वटन संभव है।

$$\begin{aligned} &\sum_{r=0}^{\infty} \frac{\lambda^r}{r!} e^{-\lambda} \\ &= e^{-\lambda} \left[1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots + \frac{\lambda^r}{r!} + \dots \right] \\ &= e^{-\lambda} \cdot e^{\lambda} \\ &= 1 \end{aligned}$$

यह स्पष्ट है कि किसी भी द्विपद वटन का पूर्ण ज्ञान हमें N और p के मानों के ज्ञान से हो जाता है क्योंकि सभी प्रायिकताएँ इन्ही दो सख्याओ से व्युत्पन्न हैं। किसी भी वंटन में ऐसे मानों को जिनमें उसकी परिभाषा होती है उम वटन के प्राचल (parameters) कहते हैं। प्वार्सों-वंटन के लिए केवल एक λ का ही मान जानना आवश्यक है। यही इस वटन का अकेला प्राचल है।

§ ७.३ वास्तविक वंटन का प्वार्सों-वंटन द्वारा सन्निकटन

अब यह देखा जा सकता है कि ऊपर जो उदाहरण दिये गये थे और जिनमें द्विपद वटन के प्रयोग में हमें हिचकिचाहट थी उनके लिए प्वार्सों-वंटन द्वारा वास्तविक प्रायिकताओं के काफी अच्छे सन्निकट (approximate) मानों के परिकलन किये जा सकते हैं। इसका कारण यह है कि सीमान्त मान की परिभाषा के अनुसार यदि N के किसी फलन $f(N)$ का सीमान्त मान g हो तो यथेष्ट रूप से बड़े N के लिए g और $f(N)$ में अंतर नग्न्य की ओर अग्रसर होता जाता है।

इस वटन का सबसे प्रसिद्ध उदाहरण है बोट-केविच (Bortkewitch) द्वारा सकलित आधार-सामग्री जिसको प्रोफेसर रोनाल्ड ए फिशर (R.A. Fisher) ने अपनी पुस्तक में भी उद्धृत किया है। इस फ्रीजी टुकड़ियों में बीस वर्षों में जो मृत्युएँ घड़े की दुल्ती के आघात से हुई थी यह उनमें मरवधित आँकड़ों पर आधारित है। इनको नीचे सारणी में दिया हुआ है।

सारणी संख्या 7.1

मृत्यु सख्या	वर्षों की बारबारता जिनमें यह मृत्यु सख्या थी
0	109
1	65
2	22
3	3
4	1
5	0
6	0

हम देखते हैं कि कुल मृत्यु-सख्या

$$(0 \times 109) + (1 \times 65) + (2 \times 22) + (3 \times 3) + (4 \times 1)$$

$$= 65 + 44 + 9 + 4$$

$$= 122 \text{ है।}$$

अर्थात् प्रति टुकड़ी प्रतिवर्ष मृत्यु संख्या ०.६१ हुई। इसलिए हम λ का मान ०.६१ ले सकते हैं और तब $e^{-\lambda} = 0.543$ (तीन दशमलव तक तक सही)। अलग अलग घटनाओं की प्रायिकता का परिकलन उस प्वासा-वटन के आधार पर जिसमें प्राचल $\lambda = 0.61$ हो नीचे दे रखा है।

सारणी संख्या 72

प्रति टुकड़ी प्रति वर्ष मृत्यु संख्या	प्रायिकता	दो सौ घटनाओं में अपेक्षित बारंबारता	वास्तविक बारंबारता
(1)	(2)	(3)	(4)
0	$e^{-\lambda} = 0.543$	108.6	109
1	$\lambda e^{-\lambda} = 0.331$	66.2	65
2	$\frac{\lambda^2}{2!} e^{-\lambda} = 0.101$	20.2	22
3	$\frac{\lambda^3}{3!} e^{-\lambda} = 0.021$	4.2	3
4	$\frac{\lambda^4}{4!} e^{-\lambda} = 0.003$	0.6	1

अपेक्षित और वास्तविक बारंबारताओं की तुलना करने से पाठकों को यह विश्वास हो जायेगा कि इस प्वासा वटन के आधार पर परिकलन करके हम वास्तविक मृत्यु संख्या का एक अच्छा सन्निकट मान प्राप्त हो सकता है। विशेष रूप से जब हम जानते हैं कि यादृच्छिक प्रयोग के फलस्वरूप बारंबारता अचर नहीं होनी और भिन्न भिन्न प्रतिदर्शों में वह भिन्न-भिन्न हो सकती है। अतः यह मान लेना असंगत नहीं समझा जा सकता कि मृत्यु संख्या एक प्वासा चर है जिसमें प्राचल का मान ०.६१ है। यद्यपि प्रकृति में यादृच्छिक चर किंग प्रकार आचरण करता है इसका ठीक पता न हमें है और

न लग सकता है तथापि प्लासो-वटर एक ऐसा सरल और सतोपजनक निरूपण है जिसके आधार पर हम घटनाओं की प्रायिकताओं का अनुमान लगा सकते हैं तथा उनके बारे में किसी हद तक भविष्यवाणी भी कर सकते हैं। यह देखा गया है कि हर एक प्रकार की आकस्मिक घटनाओं के लिए यह वटन एक अच्छे प्रतिरूप का काम देता है। यह वैसे भी स्पष्ट है क्योंकि यह द्विपद-वटन का सीमान्त रूप है जब प्राथमिक घटना की प्रायिकता p शून्यप्राय हो जाती है। प्रायिकता का शून्यप्राय होना अथवा घटना का आकस्मिक होना एक ही बात के दो रूप हैं। घटना को आकस्मिक उस समय कहते हैं जब इसकी आशा नहीं की जाती। आशा न करने का कारण यह होता है कि उस घटना की प्रायिकता बहुत कम होती है और हमारे अनुभव में ऐसी घटना के बार-बार होने की सम्भावना भी बहुत कम रहती है।

§ ७४ प्लासो-वटन के कुछ गुण

आइये, अब हम प्लासो-वटन के बारे में कुछ और जानकारी प्राप्त करें।

(१) यह वटन भी असतत है और प्लासो-वटर सभी पूर्ण-संख्याओं के बराबर मान धारण कर सकता है तथा अन्य कोई मान नहीं धारण करता।

(२) परिभाषा के अनुसार इस वटन का माध्य

$$\begin{aligned} \mu(n) = E(n) &= \sum_{n=0}^{\infty} n e^{-\lambda} \frac{\lambda^n}{n!} \\ &= \lambda \sum_{n=1}^{\infty} e^{-\lambda} \frac{\lambda^{n-1}}{(n-1)!} \end{aligned}$$

यदि हम $(n-1)$ को n से सूचित कर तो

$$\begin{aligned} \mu(n) &= \lambda \sum_{n=1}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \\ &= \lambda e^{-\lambda} \left[1 + \sum_{n=1}^{\infty} \frac{\lambda^n}{n!} \right] \\ &= \lambda e^{-\lambda} e^{\lambda} \\ &= \lambda \end{aligned}$$

(72)

इस प्रकार इस वटन का माध्य इसने प्राचल λ के बराबर होता है।

(३) इस वटन का प्रसरण (variance)

$$\begin{aligned}\sigma^2(n) &= E(n^2) - E^2(n) \\ &= \sum_{n=0}^{\infty} n^2 e^{-\lambda} \frac{\lambda^n}{n!} - \lambda^2 \\ &= \sum_{n=0}^{\infty} \{n(n-1) + n\} e^{-\lambda} \frac{\lambda^n}{n!} - \lambda^2 \\ &= \lambda^2 \sum_{n=2}^{\infty} e^{-\lambda} \frac{\lambda^{n-2}}{(n-2)!} + \lambda \sum_{n=1}^{\infty} e^{-\lambda} \frac{\lambda^{n-1}}{(n-1)!} - \lambda^2\end{aligned}$$

लेकिन $\sum_{n=2}^{\infty} \frac{\lambda^{n-2}}{(n-2)!} = e^{\lambda}$ और $\sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} = e^{\lambda}$

$$\begin{aligned}\therefore \sigma^2(n) &= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda e^{-\lambda} e^{\lambda} - \lambda^2 \\ &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda\end{aligned}$$

इस प्रकार यह एक ध्यान देने योग्य गुण है कि इस वटन का माध्य और प्रसरण दोनों ही इसके प्राचल λ के बराबर होते हैं। इस माध्य और प्रसरण का कलन हम दूसरे ढंग से भी कर सकते हैं। हमें यह तो याद ही है कि यह उस प्रकार के द्विपद-वटनों का सीमान्त रूप है जिनमें N और p का गुणनफल λ के बराबर है। द्विपद वटन में माध्य का मान Np और प्रसरण का मान Npq होता है। इसलिए हम आशा करते हैं कि प्वासो-वटन के माध्य और प्रसरण क्रमशः Np और Npq के सीमान्त मान होंगे।

$$\begin{aligned}\text{लेकिन} \quad Np &= \lambda \\ \text{और} \quad q &= 1 - p = 1 - \frac{\lambda}{N}\end{aligned}$$

λ एक अचर है, इसलिए जैसे-जैसे N का मान बढ़ता जाता है $\frac{\lambda}{N}$ का मान शून्य की ओर तथा $1 - \frac{\lambda}{N}$ का मान 1 की ओर अग्रसर होता जाता है। इस प्रकार q का सीमान्त मान 1 है।

इसलिए Npq का सीमान्त मान $\lambda \times 1 = \lambda$ है।

(४) यदि दो स्वतन्त्र प्लासो-चर हों जिनके प्राचल क्रमशः λ_1 और λ_2 हों तो इन दोनों चरों का योग भी एक प्लासो-चर है जिसका प्राचल $(\lambda_1 + \lambda_2)$ है।

ऊपर लिखे सिद्धान्त को हम एक उदाहरण द्वारा समझने की चेष्टा करेंगे। मान लीजिए एक मिल को फौज के लिए सूटों का कपड़ा बनाने का ठेका दिया जाता है। एक सूट में एक पतलून और एक कमीज है जिसके लिए कपड़ा मिल के दो विभिन्न विभागों में बनता है। बने हुए सूट में दोपों की संख्या एक यादृच्छिक-चर है जिसका वटन प्लासो-वटन माना जा सकता है। यदि पतलून में दोपों की संख्या एक प्लासो-चर हो जिसका प्राचल λ_1 है और कमीज के दोपों की संख्या भी एक प्लासो-चर हो जिसका प्राचल λ_2 है तो पूरे सूट में दोपों की संख्या अर्थात् इन दोनों दोप-संख्याओं का योग भी एक प्लासो-चर होगा और उसका प्राचल $(\lambda_1 + \lambda_2)$ होगा।

सूट के कपड़ों को छोटे-छोटे लासो वर्गों में बाँटा जा सकता है और किसी विशेष वर्ग में दोप के पाये जाने की प्रायिकता बहुत कम है। इसलिए दोपयुक्त वर्गों की संख्या के लिए प्लासो-वटन का उपयोग इस स्थिति में युक्ति-युक्त है। इन्हीं कारणों से पूरे सूट के दोपों के लिए प्लासो-वटन का उपयोग भी युक्ति-युक्त ठहराया जा सकता है। क्योंकि λ_1 से औसतन एक पतलून में पायी जानेवाली दोपसंख्या और λ_2 से औसतन एक कमीज में पायी जानेवाली दोपसंख्या सूचित होती है। इस कारण एक सूट में औसतन $(\lambda_1 + \lambda_2)$ दोपों की आँका की जा सकती है। यही कुल दोपसंख्या का प्राचल है।

ऊपर की अस्पष्ट युक्ति से हम जिस सिद्धान्त पर पहुँचते हैं उसकी सतोपजनक पथारोति उपपत्ति नीचे दी जा रही है।

मान लीजिए X और Y से दो स्वतन्त्र प्लासो-चरों को सूचित किया जाता है जिनके प्राचल λ_1 और λ_2 हैं। हम मालूम करना चाहेंगे कि यादृच्छिक-चर $(X+Y)$ का वटन क्या है। X और Y दोनों केवल पूर्ण-संख्यक मान ही धारण करते हैं। इसलिए यह स्पष्ट है कि $(X+Y)$ भी केवल पूर्ण-संख्यक मान ही धारण कर सकता है। आइए, देखें कि $(X+Y)$ के मान n धारण करने की प्रायिकता क्या है जहाँ n एक पूर्ण संख्या है। यह मान निम्नलिखित स्थितियों में धारण किया जा सकता है।

- | | | |
|----|----------|-------|
| 1. | $X=n,$ | $Y=0$ |
| 2. | $X=n-1,$ | $Y=1$ |
| 3. | $X=n-2,$ | $Y=2$ |

$$\begin{array}{lll} n & X=Y & Y=n-1 \quad \dots\dots\dots \\ n+1 & X=0 & Y=n \end{array}$$

इनमें से प्रत्येक घटना दो घटनाओं का प्रतिच्छेद है। और क्याकि ये दो घटनाएँ स्वतन्त्र हैं इसलिए इस प्रतिच्छेद की प्रायिकता इन दो घटनाओं की प्रायिकताओं का गुणनफल है। इस कारण इन ऊपर लिखी घटनाओं की प्रायिकताएँ तमस निम्नलिखित हैं—

$$1 \quad e^{-\lambda_1} \frac{\lambda_1^n}{n!} \times e^{-\lambda_2} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \lambda_1^n$$

$$2 \quad e^{-\lambda_1} \frac{\lambda_1^{n-1}}{(n-1)!} \times e^{-\lambda_2} \frac{\lambda_2}{1!} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \binom{n}{1} \lambda_1^{n-1} \lambda_2$$

$$3 \quad e^{-\lambda_1} \frac{\lambda_1^{n-2}}{(n-2)!} \times e^{-\lambda_2} \frac{\lambda_2^2}{2!} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \binom{n}{2} \lambda_1^{n-2} \lambda_2^2$$

$$n \quad e^{-\lambda_1} \frac{\lambda_1}{1!} \times e^{-\lambda_2} \frac{\lambda_2^{n-1}}{(n-1)!} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \binom{n}{n-1} \lambda_1 \lambda_2^{n-1}$$

$$n+1 \quad e^{-\lambda_1} e^{-\lambda_2} \frac{\lambda_2^n}{n!} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \binom{n}{n} \lambda_2^n$$

इसलिए $(X+Y)$ के मान n धारण करने की कुल प्रायिकता

$$P[(X+Y)=n] = \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \left\{ \lambda_1^n + \binom{n}{1} \lambda_1^{n-1} \lambda_2 + \binom{n}{2} \lambda_2^2 \right\} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!} (\lambda_1+\lambda_2)^n$$

लेकिन यदि $(X+Y)$ एक प्थामा चर होता जिसका प्राचल $(\lambda_1+\lambda_2)$ होता है, उसके मान n धारण करने की प्रायिकता भी $\frac{e^{-(\lambda_1+\lambda_2)}}{n!} (\lambda_1+\lambda_2)^n$ ही होती।

इससे यह सिद्ध हुआ कि दो स्वतन्त्र प्लासो-चरो का योग भी एक प्लासो-चर होता है और उसका प्राचल इन स्वतन्त्र प्राचलो का योग होता है। इसी प्रकार आगमिक विधि (inductive method) से यह सिद्ध किया जा सकता है कि r स्वतन्त्र प्लासो-चरो का योग भी एक प्लासो-चर होता है जिसका प्राचल इन प्लासो-चरो के प्राचलो का योग होता है। यह उपपत्ति इतनी सरल है कि उसको यहाँ देना आवश्यक नहीं समझा गया है।

§ ७'५ उदाहरण

आइए हम उस उदाहरण पर पुन विचार करें जिससे हमने प्लासो-वटन का परिचय कराया था। इसमें एक प्रार्थी को टाइपिस्ट का स्थान देने के लिए परीक्षा लेनी थी। यदि मैनेजर उन सब पृष्ठों को फिर से टकित करवाता है जिनमें एक भी दोष हो तो द्विपद टकन का उपयोग करना होगा जैसा हम पिछले अध्याय में लिख चुके हैं। परन्तु हो सकता है कि मैनेजर ऐसा न करके केवल दोषों को ठीक कर दे। ऐसी दशा में वह उन पृष्ठों की गणना नहीं करेगा जिन पर कम से कम एक दोष है परन्तु कुल दोषों की मख्या जानना चाहेगा। यदि यह मख्या बहुत अधिक हो तो दोषों के सुधारने पर पृष्ठ गदे और भड़े लगने लगेंगे। इसकी चेष्टा ऐसा टाइपिस्ट नियुक्त करने की होगी जिसके लिए इन दोषों का औसत बहुत कम हो। पहले जो टाइपिस्ट था औसतन दो पृष्ठों पर तीन गलतियाँ करता था, यदि प्रार्थी इतनी या इससे कम गलतियाँ करता है तो उसकी नियुक्ति के लिए मैनेजर को कुछ भी आपत्ति नहीं होगी।

अब भी प्रार्थी को वही परीक्षा देने के लिए कहा जाता है जिसका पिछले अध्याय में वर्णन किया जा चुका है अर्थात् उससे चार पृष्ठ टकित करने के लिए कहा जाता है और मैनेजर गलतियों को गिनता है। यदि वे ६ से कम हों तो इस प्रतिदर्श में गलतियों का मख्या औसतन पिछले टाइपिस्टों के औसत से कम है और इस प्रयोग के आधार पर प्रार्थी के अस्वीकृत करने का कोई कारण नहीं दीखता। इसके विपरीत यदि त्रुटियों की मख्या १० हो तो यद्यपि इस प्रतिदर्श में औसत पिछले टाइपिस्ट के औसत से अधिक है तथापि प्रार्थी को अस्वीकार करने के पूर्व हम यह जानना चाहेंगे कि यदि इस प्रार्थी का औसत भी १५ त्रुटि प्रति पृष्ठ होता तो इस चार पृष्ठ के प्रतिदर्श में १० त्रुटियाँ पाये जाने या इससे अधिक त्रुटियाँ पाये जाने की प्रायिकता क्या है। यदि यह प्रायिकता बहुत कम न हो तो एक न्यायशील मैनेजर प्रार्थना को एकदम अस्वीकृत न करके उसको कुछ और पृष्ठ टकित करने को देगा।

आइए, हम चार दक्कित पृष्ठों में दस या उससे भी अधिक गलतियाँ होने की प्रायिकता का बलन करें —

$$\begin{aligned}
 & P \text{ (दस अथवा उससे भी अधिक गलतियाँ)} \\
 = & 1 - P \text{ (नी या उससे कम गलतियाँ)} \\
 = & 1 - [P \text{ (शून्य गलतियाँ)} + P \text{ (एक गलती)} + P \text{ (दो गलतियाँ)} \dots \\
 & \quad \quad \quad + P \text{ (आठ गलतियाँ)} + P \text{ (नौ गलतियाँ)}] \\
 = & 1 - e^{-\lambda} \left\{ 1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots + \frac{\lambda^9}{9!} \right\} \\
 = & 1 - 0.916064 \\
 = & 0.083936
 \end{aligned}$$

§ ७६ प्वासो-वटन की सारणी

जैसे द्विपद वटन के असह्य उपयोग हैं उसी प्रकार प्वासो-वटन के भी बहुत से उपयोग हैं। अनेक मनुष्यों के बार-बार एक ही प्रकार के परिकलन करने की वृथा मेहनत को बचाने के लिए सारणियाँ तैयार कर ली गयी हैं। इन सारणियों में λ के विभिन्न मानों के लिए प्वासो चर के 0, 1, 2, 3, आदि मान धारण करने की प्रायिकताएँ दे रखी हैं। कुछ और भी सारणियाँ हैं जिनमें प्वासो-चरों की सखी आपेक्षिक बारबारताएँ दी हुई हैं। जब किसी को प्रायिकताओं के कलन के लिए अथवा परिकल्पना की परीक्षा के लिए प्वासो-वटन का उपयोग करना होता है तब सब परिकलन नये सिरे से नहीं करने पड़ते। उसे विशेष λ के मान के लिए सारणी को देखना ही सफेद होता है।

नीचे इस प्रकार की सारणी का एक नमूना दे रखा है। जिस सारणी का ऊपर के उदाहरण में प्रयोग हुआ है वही यहाँ दे रखी है। यह ध्यान देने योग्य बात है कि द्विपद वटन की तरह प्वासो-वटन भी असतत है। इस प्रकार कोई भी सख्या r ऐसी नहीं है जिससे अधिक चर का मान होने की प्रायिकता ठीक पाँच प्रतिशत या ठीक एक प्रतिशत हो। परन्तु एक छोटी-से छोटी पूर्ण-सख्या मालूम की जा सकती जिससे अधिक मान धारण करने की प्रायिकता पाँच प्रतिशत से कम हो। यदि हम यह निश्चय कर लें कि किसी परिकल्पना के आधार पर प्रेषित सख्या के बराबर अथवा उससे अधिक मान धारण करने की प्रायिकता पाँच प्रतिशत से कम होने पर हम उस परिकल्पना को अस्वीकृत कर देंगे तो हम प्रयोग से पहले ही एक ऐसी सख्या निश्चित

कर सकते हैं कि प्रयोग का फल उससे अधिक होने पर हम परिकल्पना को झूठी समझेंगे।

सारणी सख्या 73

प्लासो बटन ($\lambda = 6$) के लिए सचयी प्रायिकता फलन $F(r)$

r (1)	$F(r)$ (2)
0	0 002468
1	0 017341
2	0 061958
3	0 151192
4	0 285045
5	0 445668
6	0 606291
7	0 743968
8	0 847226
9	0 916064

r (1)	$F(r)$ (2)
10	0 957367
11	0 979897
12	0 991161
13	0 996360
14	0 998588
15	0 999479
16	0 999813
17	0 999931
18	0 999970
19	0 999982

विस्तृत सारणी के लिए देखिए
'Molna's Tables'



अध्याय ८

प्रसामान्य वंटन (Normal Distribution)

§ ८.१ गणतीय वंटनो का महत्त्व

अभी तक हमने द्विपद और प्वासो-वंटनो का अध्ययन किया है जो असतत हैं और केवल पूर्ण-संख्या मान धारण करते हैं। परन्तु हम जानते हैं कि कुछ यादृच्छिक चर ऐसे भी होते हैं जो दो सीमान्त मानों के बीच के सभी मानों को धारण कर सकते हैं। ऐसे चरों का एक उदाहरण मनुष्य की ऊँचाई है। इस प्रकार के चरों का एक घनत्व-फलन (*density function*) होता है। जंसा हम पहिले ही देख चुके हैं, किसी भी विशेष मान को धारण करने की प्रायिकता इस चर के लिए शून्य होती है। परन्तु किसी अल्पतम अंतराल में भी इस चर के स्थित होने की प्रायिकता शून्य से भिन्न हो सकती है। इस प्रायिकता को अंतराल की लम्बाई से विभाजित करने से हमें इस अंतराल में प्रायिकता का घनत्व मालूम होता है। जैसे-जैसे अंतराल छोटा होता जाता है सतत वंटनो में यह घनत्व एक विशेष मन्थ्या की ओर अग्रसर होता जाता है। जो सरया इन घनत्व का सीमान्त रूप है वही उस अंतराल के मध्य बिंदु पर वंटन का घनत्व माना जाता है। घनत्व फलन चर के मान और उस मान से सगत घनत्व के मन्थ्य की प्रदर्शन करता है।

मान लीजिए कि X एक ऐसा सतत चर है और उसका घनत्व फलन $f(x)$ है। यदि इस चर की समष्टि में से हम एक प्रतिदर्श का चयन करें जिसका परिमाण n हो तो प्रश्न उठता है कि इस प्रतिदर्श के माध्य का क्या वंटन होगा। यदि इस चर के n मानों को जो प्रतिदर्श में विद्यमान हैं, हम $x_1, x_2, x_3, \dots, x_{n-1}, x_n$ से सूचित करें तो हमें प्रायिकता $P \left[\frac{x_1 + x_2 + \dots + x_n}{n} \leq k \right]$ का परिवर्तन k के विभिन्न मानों के लिए करना है। इस प्रायिकता को हम निम्नलिखित बहुल समाकल (multiple integral) से सूचित करते हैं।

$$P \left[\sum_{r=1}^n \tau_r \leq nk \right] = \int_{-\infty}^{nk} \int_{-\infty}^{nk-x_1} \int_{-\infty}^{nk-(x_1+x_2)} \dots \int_{-\infty}^{nk-\sum_{i=1}^{n-1} x_i} f(x_1) f(x_2) \dots f(x_n) dx_1 dx_2 \dots dx_n \quad (81)$$

साधारणतया इस समाकल का मूल्यांकन करना यदि असंभव नहीं तो बहुत कठिन अवश्य होता है। लेकिन जैसा हम पहिले कई बार कह चुके हैं, सांख्यिकी में प्राथमिकताओं के एकदम यथार्थ मान जानना आवश्यक नहीं है। सन्निकट मान (approximate value) ही यथेष्ट होता है। आपकी कही यह तो मदेह नहीं हो रहा है कि सांख्यिकी का आधार बहुत कमजोर है—इसमें कुछ भी तथ्य नहीं है और सभी सन्निकटन मान हैं? अनुमान और सन्निकट माप का तो हर एक विज्ञान में और साधारण दिनचर्या में जगह जगह प्रयोग किया ही जाता है। यह सत्य है कि वैज्ञानिकों ने यथार्थतम मापों के लिए ऐसे-ऐसे यंत्रों का आविष्कार किया है कि उनकी तारीफ बिये बिना नहीं रहा जाता। परन्तु कोई भी वैज्ञानिक यह दावा नहीं करता कि ये माप विलकुल यथार्थ हैं।

मान लीजिए कि कोई मनुष्य एक लोहे की छड़ की लंबाई माप रहा है। यदि उसकी यथार्थ माप करने की जिद है तो यह कार्य असंभव होगा। नापना तो दूर, पहिले इस लंबाई की परिभाषा देना ही असंभव होगा। हम जानते हैं कि छड़ अणुओं की बनी होती है। ये अणु अस्थिर होते हैं और इनमें बराबर कंपन (vibration) होता रहता है। यथार्थता के लिए छड़ की लंबाई किमी विशेष धन और विशेष रेखा से सवधित होगी। परन्तु क्या कभी ऐसा किया जाता है? व्यावहारिक रूप से इस लंबाई में कोई विशेष अंतर नहीं दिखाई देता, यदि उसको गर्म या ठंडा न किया जाए। इस कारण हम इन मामूली परिवर्तनों की पर्वाह नहीं करते और एक सन्निकटतम लंबाई मालूम करते हैं। मान लीजिए, इस आणविक कंपनों के कारण एक छड़ की लंबाई 10 123255 सेंटीमीटर से 10 123256 सेंटीमीटर के बीच विभिन्न मानों को धारण करती रहती है। इस मामूली से अंतर को आसानी से भुलाया जा सकता है। व्यावहारिक जीवन में प्रायः एक प्रतिशत की यथार्थता (precision) यथेष्ट समझी जाती है। एक प्रतिशत की यथार्थता से हमारा तात्पर्य यह है कि यदि यथार्थ माप एक सौ है तो सन्निकट माप निम्नानवे और एक सौ एक के बीच की ही कोई संख्या

हो। भौतिकी अथवा रसायन में हमारा लक्ष्य एक प्रति दस हजार की यथार्थता हो सकता है, परन्तु प्रत्येक अवस्था में यथार्थता की भी कोई सीमा होती है जहाँ खना ही पड़ता है।

सांख्यिकी में हम वास्तविक घटनों का सन्निकटन कुछ गणितीय वटनों (mathematical distributions) के द्वारा करते हैं। यह सन्निकट वटन ऐसा होना चाहिए कि इसके ओर वास्तविक वटन के सचयी-वार-वारता-वटनों में कोई बिचोप अंतर न हो। कितने अंतर तक को सहन किया जा सकता है यह व्यक्तिगत रुचि और जरूरत पर निर्भर है। इस प्रकार के सन्निकटन से अमीमित लाभ हैं। इस गणितीय वटन के माध्य, प्रसरण और अन्य घूर्णों का परिकलन अपेक्षाकृत सरल होता है। इसके अन्य गुणों की व्याख्या भी बड़ी आसानी से की जा सकती है। कुछ गणितीय वटनों का सन्निकट वटनों के रूप में विभिन्न परिस्थितियों में विभिन्न व्यक्तियों द्वारा प्रयोग किया जा सकता है। ऐसा हम द्विपद-वटन और प्वासो वटन के लिए पहिले ही देख चुके हैं। ऐसे वटनों के लिए सारणी तैयार कर ली जाती है और जब कभी भी सन्निकट वटन का उपयोग किया जाता है, इस सारणी को देखकर प्रायिकताओं का परिकलन किया जाता है। इस सारणी को देखकर प्रायिकताओं का परिकलन अथवा परिकल्पनाओं के धारे में फैसला किया जा सकता है। यदि ऐसा न किया जाय तो दो ही बातें हो सकती हैं— या तो जिस चर का अध्ययन किया जा रहा है उसके वास्तविक वटन का किसी को ज्ञान नहीं है। ऐसी अवस्था में यदि वह किसी सन्निकटन का उपयोग नहीं करना चाहता जो उसे चर के बारे में किसी भी निश्चय पर पहुँचने का विचार छोड़ देना चाहिए। यदि वास्तविक वटन ज्ञात भी हो तो चर के विभिन्न मानों के लिए प्रायिकताओं का परिकलन या वटन के प्रतिगता-बिंदुओं (percentage points) का मालूम करना बहुत ही कठिन हो जायगा। यही नहीं बल्कि इस कठिनाई का सामना बार-बार हर नयी स्थिति के लिए करना होगा। इस बात की संभावना बहुत कम है कि किसी भी वास्तविक वटन का प्रयोग दुबारा करने की आवश्यकता पड़े।

§ ८२ प्रसामान्य वटन की परिभाषा

सांख्यिकी ने एक बड़ी आश्चर्यजनक और महत्वपूर्ण खोज की है। उन्होंने यह सिद्ध कर दिया है कि किसी चर का वास्तविक वटन चाहे कुछ भी हो, परन्तु उसके एक बड़े प्रतिदर्श के माध्य का सन्निकटन एक सतत यादृच्छिक चर द्वारा किया जा सकता है। इस सतत चर का प्रायिकता घनत्व-फलन $\phi(\bar{x})$ यह है—

$$\phi(\bar{x}) = \frac{1}{\sqrt{2\pi\sigma^2/\sqrt{n}}} e^{-\frac{1}{2} \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right)^2} \dots \dots \dots (8.2)$$

जहाँ μ और σ^2 क्रमशः मूल वटन के माध्य और प्रसरण है, n प्रतिदर्श परिमाण है, π एक वृत्त की परिधि और उसके व्यास का अनुपात है एवं e की परिभाषा वही है जो हम पहिले ही प्वासॉन्-वटन पर विचार करते समय दे चुके हैं।

जिन चरों के वटन का रूप ऊपर लिखित वटन के प्रकार का होता है वे प्रसामान्य चर (Normal variates) कहलाते हैं और तत्समवधी वटनों को प्रसामान्य वटन (Normal distribution) कहते हैं। यह आप देख ही सकते हैं कि μ और σ/\sqrt{n} के विभिन्न मानों के लिए हमें विभिन्न प्रसामान्य वटन प्राप्त होते हैं। इस कारण ये ही प्रसामान्य वटन के प्राचल (parameters) हैं। ये प्राचल क्रमशः प्रसामान्य वटन के माध्य और मानक विचलन भी हैं। प्रतिदर्श-परिमाण तो प्रसामान्य वटन के परिचय में प्रासंगिक मात्र था और प्रसामान्य चर की परिभाषा में इसका कोई स्थान नहीं है। प्रसामान्य चर के घनत्व-फलन को हम

$$\phi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2} \dots \dots \dots (8.3)$$

से सूचित करते हैं जहाँ μ और σ क्रमशः इस चर के माध्य और मानक विचलन हैं।

§ ८.३ प्रसामान्य वंटन के कुछ महत्त्वपूर्ण गुण

प्रसामान्य वटन का उपयोग समझने से पहिले हमें उसके कुछ गुणों से परिचित हो जाना चाहिए।

(१) यदि X_1 , और X_2 दो स्वतंत्र प्रसामान्य चर हों जिनके प्राचल (μ_1, σ_1) और (μ_2, σ_2) हैं तो इन दोनों चरों का योग $(X_1 + X_2)$ भी एक प्रसामान्य चर है जिसके प्राचल $(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$ होते हैं।

(२) ऊपर लिखित फल को आगमिक विधि से किन्हीं भी N प्रसामान्य चरों पर लागू किया जा सकता है। यदि इन N चरों के प्राचल क्रमशः $(\mu_1, \sigma_1), (\mu_2, \sigma_2), \dots, (\mu_n, \sigma_n), \dots, (\mu_N, \sigma_N)$ हों और यदि वे चर स्वतंत्र हों तो इनका योग भी एक प्रसामान्य चर होता है जिसके प्राचल $\left(\sum_{i=1}^N \mu_i, \sqrt{\sum_{i=1}^N \sigma_i^2} \right)$ हैं।

(३) यदि प्रसामान्य चर X का माध्य μ और प्रसरण σ^2 है तो उसका कोई भी एक-घात फंक्शन (linear function) $aX+b$ भी एक प्रसामान्य चर है जिसके माध्य और प्रसरण क्रमशः $a\mu+b$ तथा $a^2\sigma^2$ हैं। इस चर के प्राचल ऊपर-लिखित होंगे यह आसानी से देखा जा सकता है, क्योंकि

$$\begin{aligned} E(aX+b) &= E(aX) + E(b) \\ &= aE(X) + b \\ &= a\mu + b \end{aligned}$$

$$\begin{aligned} \text{इसी प्रकार } V(aX+b) &= V(aX) \\ &= a^2 V(x) \\ &= a^2 \sigma^2 \end{aligned}$$

जब हम कहते हैं कि किसी यादृच्छिक चर का घनत्व फंक्शन $f(x)$ है तो इसका अर्थ यह होता है कि यदि dx छोटा हो तो x और $x+dx$ के बीच इस चर के मान के पाये जाने की प्रायिकता लगभग $f(x) dx$ होती है। इस तरह

$$P[x' < X < x' + dx'] = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx'$$

$$\begin{aligned} \therefore P[x' < aX+b < x'+dx] &= P\left[\frac{x-b}{a} < X < \frac{x'-b}{a} + \frac{dx'}{a}\right] \\ &= \frac{1}{\sigma\sqrt{2\pi}} \left[e^{-\frac{1}{2}\left(\frac{x-b}{\sigma}-\mu\right)^2/\sigma^2} \right] \frac{dx'}{a} \\ &= \frac{1}{a\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-(a\mu+b)}{a\sigma}\right]^2} dx' \end{aligned} \quad (8.4)$$

यानी $(aX+b)$ एक प्रसामान्य चर है जिसके प्राचल $(a\mu+b, a\sigma)$ हैं।

(४) यदि $a = \frac{1}{\sigma}$ और $b = -\frac{\mu}{\sigma}$ हो तो $\frac{x-\mu}{\sigma}$ का घनत्व फंक्शन निम्न-लिखित होगा।

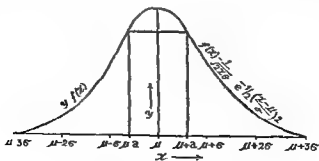
$$\phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (8.5)$$

यह एक प्रसामान्य चर का घनत्व-फल है जिसका माध्य शून्य तथा प्रसरण एक है। इस चर के वंटन को मानकित प्रसामान्य वंटन (standardised Normal distribution) कहते हैं। इसको $N(0,1)$ से सूचित किया जाता है और इसे "प्रसामान्य शून्य एक" पढ़ते हैं। इसी प्रकार जिस प्रसामान्य वंटन का माध्य μ तथा मानक विचलन σ हो उसे $N(\mu, \sigma)$ से सूचित किया जाता है।

(५) ऊपर दिये हुए गुण से यह पता चलता है कि यदि इस मानकित प्रसामान्य वंटन के प्रतिशतता-बिन्दुओं की सारणी तैयार की जाय तो आसानी से किसी भी प्रसामान्य वंटन $N(\mu, \sigma)$ के प्रतिशतता बिन्दुओं का कलन किया जा सकता है। इस प्रकार की सारणी सांख्यिकों ने तैयार कर रखी है।

मान लीजिए, हमें किसी प्रसामान्य वंटन का प्रसरण σ^2 ज्ञात है और हम इस परिकल्पना की जाँच करना चाहते हैं कि वंटन का माध्य μ है। हम n परिमाण का एक प्रतिदर्श (sample) लेकर प्रतिदर्श माध्य \bar{x} का परिकलन कर सकते हैं। यदि परिकल्पना सत्य है तो $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ एक $N(0,1)$ चर है। इस कारण हम सारणी द्वारा $P\left[N(0,1) \geq \frac{|\bar{x}-\mu|}{\sigma/\sqrt{n}}\right]$ मालूम कर सकते हैं। यदि यह प्रायिकता बहुत कम हो तो हमारा परिकल्पना पर सदेह होना और इस कारण उसे अस्वीकार कर देना स्वाभाविक है।

(६) यदि हम प्रसामान्य चर X के मान और उसके घनत्व फलन के बीच एक ग्राफ खींचें तो उसकी नकल इस प्रकार की होगी जैसी नीचे के चित्र में दिखायी गयी है।

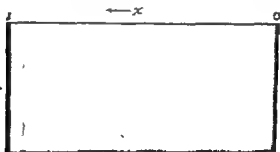


चित्र २५— $N(\mu, \sigma)$ का घनत्व-फल

ऐसा मालूम होता है कि किसी घटी को उलट कर रख दिया हो। माध्य के दोनों ओर का वटन एक-सा होता है। जो प्रायिकता घनत्व $(\mu+a)$ पर होता है वही $(\mu-a)$ पर भी होता है। इस वटन का बहुलक (mode) और माध्य बराबर होते हैं। यह चर छोटे-से-छोटे और बड़े-से-बड़े हर एक मान को धारण करता है, परंतु जैसे-जैसे मान माध्य से दूर होता जाता है, उसका प्रायिकता-घनत्व कम होता जाता है और शून्य की ओर अग्रसर होता जाता है।

§ ८.४ प्रसामान्य वटन द्विपद वटन का एक सीमान्त रूप

इससे पहिले कि हम परिकल्पना की जाँच में प्रसामान्य चर के उपयोग का अध्ययन करें आप शायद यह जानना चाहेंगे कि किसी भी वटन के लिए प्रतिदर्श-माध्य प्रसामान्य चर की ओर कैसे अग्रसर होता है। हम एक ऐसे द्विपद वटन के उदाहरण से जिसमें $p = \frac{1}{2}$ हो, इसे समझने की चेष्टा करेंगे। मान लीजिए कि हम एक सिक्के को उछालते हैं। इस यादृच्छिक प्रयोग के दो ही फल हो सकते हैं, चित या पट। यदि हम एक यादृच्छिक चर की ऐसी परिभाषा करें कि वह चित आने पर 1 और पट आने पर 0 मान को ग्रहण करता है तो इस वटन का दंड-चित्र (bar diagram) नीचे चित्र सब्या २६ के समान होगा।



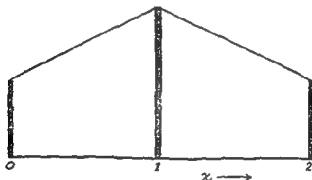
चित्र २६—द्विपद $(1, \frac{1}{2})$ का दंडचित्र

इस वटन का माध्य $\frac{1}{2}$ तथा मानक विचलन भी $\frac{1}{2}$ है

$$\begin{aligned} \text{क्योंकि } \mu &= E(X) = 0 \times \frac{1}{4} + 1 \times \frac{1}{4} \\ &= \frac{1}{4} \end{aligned}$$

$$\begin{aligned} \sigma^2 &= E(X^2) - E^2(X) \\ &= [0^2 \times \frac{1}{2} + 1^2 \times \frac{1}{2}] - (\frac{1}{2})^2 \\ &= \frac{1}{2} - \frac{1}{4} \\ &= \frac{1}{4} \\ \therefore \sigma &= \frac{1}{2} \end{aligned}$$

यदि सिक्का दो बार उछाला जाय और इन दो प्रयोगों से संबंधित चरों के माध्य का परिवर्तन किया जाय तो वह तीन मान धारण कर सकता है—0, $\frac{1}{2}$ और 1 और इनको ग्रहण करने की प्रायिकताएँ क्रमशः $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$ हैं। इसका दृढ़-चित्र चित्र संख्या २७ में दिखाया गया है।

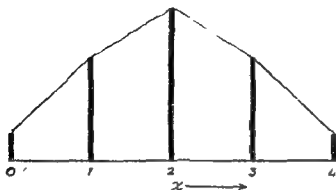
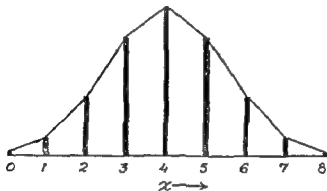
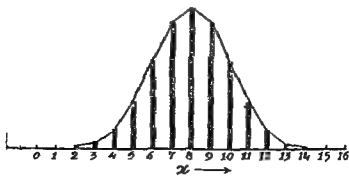


चित्र २७—द्विपद $(2, \frac{1}{2})$ का बंधचित्र

इनके माध्य और प्रसरण क्रमशः $\frac{1}{2}$ और $\frac{1}{4}$ हैं।

प्रतिदर्श-परिमाण चार होने पर प्रतिदर्श माध्य पाँच मानों 0, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$ तथा 1 को क्रमशः $(\frac{1}{4})^4$, $4(\frac{1}{4})^4$, $6(\frac{1}{4})^4$, $4(\frac{1}{4})^4$ तथा $(\frac{1}{4})^4$ की प्रायिकता के साथ ग्रहण करता है। इस माध्य के वटन के माध्य तथा प्रसरण क्रमशः $\frac{1}{2}$ तथा $\frac{1}{16}$ हैं। इसका दृढ़-चित्र चित्र संख्या २८ में दिखाया गया है।

प्रतिदर्श-परिमाण 8 और 16 से संबंधित दृढ़-चित्र भी पृ० १३६ दिये हुए हैं। (२९, ३० चित्र)। इन सभी चित्रों में (पहिले को छोड़कर) माध्य पर की प्रायिकता को,

चित्र २८—द्विपद $(4, \frac{1}{2})$ का दंडचित्रचित्र २९—द्विपद $(6, \frac{1}{2})$ का दंडचित्रचित्र ३०—द्विपद $(16, \frac{1}{2})$ का दंडचित्र

जो अन्य सब प्रायिकताओं से अधिक है, एक चार सेंटीमीटर ऊंची रेखा से सूचित किया गया है, यद्यपि विभिन्न प्रतिदर्श-परिमाणों के लिए इस मान $\frac{1}{2}$ को ग्रहण करने की प्रायिकताएँ अलग-अलग हैं। आपने यह देखा होगा कि जैसे-जैसे प्रतिदर्श परिमाण बढ़ता जाता है जैसे-जैसे दृष्ट-चित्र के दृष्ट एक दूसरे के पास आते जाते हैं। यदि इन दृष्टों के सिरो को मिलाती हुई एक वक्र रेखा खींची जाय तो जैसे-जैसे प्रतिदर्श-परिमाण बढ़ता जाता है जैसे-जैसे इन वक्र की शकल घटीनुमा वक्र की जैसी होती जाती है।

इससे भी अच्छी तुलना दो दृष्टों के बीच के मानों की तत्समवधी सघयी प्रायिकताओं से हो सकती है जो इन द्विपद वटनों और प्रसामान्य वटनों के आधारपर परिकल्पित की जाये जिनके गाध्य और प्रसरण द्विपद वटन के गाध्य और प्रसरण के बराबर हों। नीचे सारणी में $\frac{n}{10}, \frac{2n}{10}, \frac{3n}{10}, \frac{4n}{10}, \frac{5n}{10}, \frac{6n}{10}, \frac{7n}{10}, \frac{8n}{10}, \frac{9n}{10}$ तथा n पर द्विपद वटन और प्रसामान्य वटन की सचयी प्रायिकताएँ दी हुई हैं।

आगे की सारणी से यह प्रत्यक्ष ज्ञात होता है कि जैसे-जैसे प्रतिदर्श परिमाण बढ़ता जाता है द्विपद-वटन का सचयी प्रायिकता-फलन अधिकाधिक प्रसामान्य वटन के सचयी बारवारता-फलन के बराबर होता जाता है। इस उदाहरण में हमने p और q को द्विपद वटन के लिए बराबर रखा था। यदि p और q में अंतर बहुत अधिक हो तो इन दोनों फलनों के बराबर होने के लिए बहुत अधिक प्रतिदर्श परिमाण की आवश्यकता होगी।

§ ८५ त्रुटियों का वटन

वैज्ञानिकों ने यह देखा है कि चाहे कितनी भी होशियारी से माप लिया जाय, माप में कुछ-न-कुछ त्रुटि रह ही जाती है।

मान लीजिए कि एक पैमाना है जिसमें एक इंच के दसवें भाग पर निशान लगे हुए हैं। यदि हम इसकी मदद से किसी वस्तु को इंच के सौवें हिस्से तक नापना चाहते हैं तो यह काम हमारे लिए इस पैमाने से करना संभव नहीं है। यदि हमारे पास कोई पैमाना नहीं हो तो हमें इंच के दूसरे दशमलव स्थान को अनुमान द्वारा प्राप्त करना होगा। यह कैसे हो सकता है कि यथार्थ अंक का ही अनुमान लगे? गलती होना अवश्यभावी और स्वाभाविक है। यद्यपि लंबाई वही बनी रहती है तो भी एक ही मनुष्य उस ही वस्तु को बार-बार नापने पर इस अंक का अलग-अलग अनुमान

सारणी सख्या 81
द्विपद और प्रसामान्य वटनो की सचयी प्रायिकतायो की तुलना

प्रतिदर्श परिमाण n	$X =$ वटन	$\frac{n}{10}$ (3)	$\frac{2n}{10}$ (4)	$\frac{3n}{10}$ (5)	$\frac{4n}{10}$ (6)	$\frac{5n}{10}$ (7)	$\frac{6n}{10}$ (8)	$\frac{7n}{10}$ (9)	$\frac{8n}{10}$ (10)	$\frac{9n}{10}$ (11)	n
1	द्विपद	5000	5000	5000	5000	5000	5000	5000	5000	5000	1 0000
	प्रसामान्य	2119	2743	3446	4207	5000	5793	6554	7357	7881	1 0000
2	द्विपद	2500	2500	2500	2500	2500	2500	2500	2500	2500	1 0000
	प्रसामान्य	1292	1977	2843	3897	5000	6103	7157	8023	8708	1 0000
4	द्विपद	0625	0625	3125	3125	6875	6875	6875	9375	9375	1 0000
	प्रसामान्य	0548	1151	2119	3446	5000	6554	7881	8849	9452	1 0000
8	द्विपद	0039	0352	1445	3633	6367	6367	8555	9648	9961	1 0000
	प्रसामान्य	0119	0455	1292	2843	5000	7157	8708	9545	9881	1 0000
16	द्विपद	0003	0106	0245	2272	5982	7728	9616	9755	9997	1 0000
	प्रसामान्य	0007	0082	0548	2119	5000	7881	9452	9918	9993	1 0000
32	द्विपद	0000	0003	0045	1077	5700	8595	9900	9997	1 0000	1 0000
	प्रसामान्य	0000	0003	0119	1292	5000	8708	9881	9997	1 0000	1 0000

लगा सकता है। यदि अनुमान लगाने की इस क्रिया को बार-बार दुहराया जाय तो वास्तविक माप और इस प्रकार अनुमानित माप के बीच के अंतर (जिसे मापत्रुटि कहा जा सकता है) का वटन किस प्रकार का होगा? अनुभव के आधार पर यह जाना गया है कि इस वटन का एक अच्छा सन्निकटित रूप प्रसामान्य वटन है।

यह देखा गया है कि यदि हम किसी भी कार्य में बहुत अधिक यथार्थता प्राप्त करने का प्रयत्न करते हैं और इसके होते हुए भी कुछ त्रुटि हो जाती है तो यह त्रुटि प्रसामान्य-चर होती है। इसका सबसे अच्छा उदाहरण किसी छोटे से निशाने पर गोली मारने का प्रयत्न है। इस उदाहरण पर पहिले भी हम किसी दूसरे प्रसंग में विचार कर चुके हैं। यहाँ हवा का जरा-सा झोका, बनावट में जरा-सा अंतर, बंदूक को साधे हुए हाथ का तनिक-सा कपन अथवा अन्य कोई भी कारण त्रुटि उत्पन्न कर सकता है। त्रुटियों के प्रसामान्य चर होने का यही कारण बताया जाता है। विभिन्न कारणों से जो त्रुटियाँ होती हैं उनके विभिन्न अटन हो सकते हैं परन्तु समस्त प्रेक्षित त्रुटियों की सख्या इन सब विभिन्न त्रुटियों की सख्याओं का योग होगी। जैसा हम द्विपद पर के लिए देख चुके हैं, यह सिद्ध किया जा सकता है कि इन अनेक चरों के योग अथवा माध्य का वटन प्रायः प्रसामान्य होगा।

विभिन्न कारणों के मंचित प्रभाव का एक कौतूहल-जनक उदाहरण एक व्यक्ति की ऊँचाई है। जन्म सबधी उपादान कारणों के अलावा, जो शायद सबसे अधिक महत्वपूर्ण है, सँकड़ों अन्य कारण व्यक्ति की ऊँचाई पर प्रभाव डालते हैं। ऊपर के तर्क के अनुसार यह आशा की जाती है कि व्यक्तियों की ऊँचाइयों का वटन प्रसामान्य होना चाहिए और प्रेक्षण द्वारा यह देखा गया है कि यदि काफी बड़े प्रतिदर्श में मनुष्यों की ऊँचाइयों का प्रेक्षण किया जाय तो मालूम होगा कि इनका वटन लगभग प्रसामान्य है।

गाउस (Gauss) ने इस वटन को पहिले त्रुटियों के वटन के रूप में ही खोजा था। इस कारण इसकी त्रुटियों का वटन (law of errors) अथवा गाउस का वटन भी कहा जाता है। आपको यह कौतूहल होना स्वामाविक है कि इस प्रकार के जटिल वटन का विचार किम प्रकार सूँठ में किसी को आया होगा। आपके इस कौतूहल को शांत करने के लिए इस वटन की सैद्धान्तिक व्युत्पत्ति की रूपरेखा हम नीचे दे रहे हैं।

§ ८-६ गाउस के त्रुटि-वटन की व्युत्पत्ति

मान लीजिए कि किसी वस्तु का वास्तविक माप μ (म्यू) है। इस वस्तु को यदि n बार नापें तो हमें विभिन्न माप x_1, x_2, \dots, x_n प्राप्त होंगे। यदि हमें माप x_r

प्राप्त होता है तो इसमें त्रुटि $(x_r - \mu)$ है। हम इस त्रुटि को z_r से सूचित करेंगे। इस तरह

$$z_1 = x_1 - \mu, \quad z_2 = x_2 - \mu, \quad z_r = (x_r - \mu), \\ z_{n-1} = x_{n-1} - \mu, \quad z_n = (x_n - \mu)$$

यदि हम त्रुटि के परास (range) को छोटे-छोटे अंतरालों में विभाजित कर दें जिन सबका परिमाण Δz हो तो माप के z और $z + \Delta z$ के बीच में पाये जाने की प्रायिकता दो अवयवों पर निर्भर करती है।

(१) अंतराल का परिमाण Δz

(२) त्रुटि का प्रायिकता घनत्व फलन जो त्रुटि विशेष z से संबंधित है। इसे हम $f(z)$ से सूचित करेंगे। हमारा उद्देश्य हम फलन $f(z)$ का पता चलाना है। हम फलन के बारे में पहिले हम दो अभिधारणाएँ (postulates) लेकर चलते हैं।

(१) z के जिस मान के लिए इन फलन का मान महत्तम हो जाता है वह है $z=0$

(२) जहाँ जहाँ z का मान बढ़ता जाता है तब-तब $f(z)$ का मान कम होता जाता है और शून्य की ओर अभिसर होता जाता है।

ये अभिधारणाएँ अनुभव पर आधारित हैं। यदि हम सावधानी से किसी वस्तु का यथार्थ माप प्राप्त करने की चेष्टा करें तो यह स्वाभाविक है कि कम त्रुटि होने की प्रायिकता अधिक और अधिक त्रुटि होने की प्रायिकता कम होगी। बहुत अधिक त्रुटि होना प्रायः असंभव है, इसलिए ऐसी घटना के लिए $f(z)$ का मान शून्यप्राय होना ही चाहिए।

यदि z और $z + \Delta z$ के बीच में प्रेक्षित माप के पाये जाने की प्रायिकता को IV से सूचित करें तो

$$IV = f(z) \Delta z \quad (8.6)$$

यदि समस्त मापों की संख्या n हो, तो z और $z + \Delta z$ के बीच के मापों की प्रत्याशित संख्या

$$nIV = n f(z) \Delta z \quad (8.7)$$

यदि ये सब त्रुटियाँ एक दूसरे से स्वतंत्र हों अर्थात् एक माप के ज्ञान से दूसरे मापों के घटने में कोई अंतर न पड़े तो इन विचलनों के मध्य (combination) की प्रायिकता L इन विभिन्न प्रायिकताओं का गुणनफल होगी।

$$L = f(z_1)f(z_2) \dots f(z_n) (\Delta z)^n \quad (8\ 8)$$

ऊपर के समीकरण में दोनों ओर का लघुगणक (logarithm) लेने पर

$$\log L = \sum_{r=1}^n \log f(z_r) + n \log(\Delta z) \quad (8\ 9)$$

लघुगणक की परिभाषा

यदि आप लघुगणक के उपयोग से परिचित नहीं हैं तो आपको यह जानने की इच्छा होगी कि लघुगणक क्या होता है।

आप सख्या e से तो परिचय प्राप्त कर ही चुके हैं। $\log L$ की परिभाषा निम्न लिखित समीकरण द्वारा दी जाती है।

$$e^{\log L} = L \quad (8\ 10)$$

$$\text{इसी प्रकार } e^{\log M} = M \quad (8\ 11)$$

ऊपर के समीकरणों का गुणा करने पर हम देखते हैं कि

$$e^{\log L + \log M} = LM \quad (8\ 12)$$

इस प्रकार दो या अधिक सख्याओं के गुणनफल का लघुगणक उनके पृथक् पृथक् लघुगणकों का योग होता है। लघुगणक के इसी गुण का ऊपर $\log L$ के परिकलन में उपयोग किया गया है।

हम निम्नलिखित प्रतिबंध (restrictions) को दृष्टि में रखते हुए फलन $f(z)$ का चुनाव करते हैं।

(१) फलन $f(z)$ प्रायिकता का घनत्व फलन है। इसलिए z के पूर्ण परास- ∞ से $+\infty$ —में $f(z)$ का समाकल (integral) अथवा विभिन्न फलनों का योग 1 होना चाहिए

$$\int_{-\infty}^{+\infty} f(z) dz = 1 \quad (8\ 13)$$

(२) इन श्रुटिया का माध्य शून्य है

$$\sum_{r=1}^n z_r = 0 \quad (8\ 14)$$

(३) L या $\log L$ इन x_1, x_2, \dots, x_n आदि मापा के माध्य के लिए महत्तम हो जाती है।

अवकल की परिभाषा—

यदि $F(x)$ कोई सतत चर हो और उसका मान $x=a$ पर महत्तम होता हो तो यह सिद्ध किया जा सकता है कि—

$$\lim_{h \rightarrow 0} \frac{F(a+h) - F(a)}{h} = 0$$

$$\text{और } \lim_{h \rightarrow 0} \frac{F(a) - F(a-h)}{h} = 0$$

$$\text{जब कभी भी } \lim_{h \rightarrow 0} \frac{F(a+h) - F(a)}{h} = \lim_{h \rightarrow 0} \frac{F(a) - F(a-h)}{h}$$

हो तो हम कहते हैं कि फलन $F(x)$ का $x=a$ पर अवकलन (differentiation) किया जा सकता है और इन अनुपातों के सीमान्त मानों को जो बराबर हैं हम $x=a$ पर $F(x)$ का अवकल (differential coefficient) कहते हैं। इसको $F'(a)$ से सूचित किया जाता है। इस प्रकार x के विभिन्न मानों के लिए विभिन्न अवकल प्राप्त किये जा सकते हैं और ये अवकल भी x के फलन समझे जा सकते हैं जिन्हें $F'(x)$ अथवा $\frac{dF(x)}{dx}$ से सूचित करते हैं।

यह सिद्ध किया जा सकता है कि $\frac{d \log f(x)}{dx} = \frac{1}{f(x)} \frac{df(x)}{dx}$, इस कारण ऊपर के समीकरण को निम्नलिखित रूप में रखा जा सकता है (8 15)

$$\sum_{r=1}^n \phi(z_r) = 0 \quad \text{जहाँ } \phi(z_r) = \frac{df(z_r)/dz_r}{f(z_r)} \quad (8\ 16)$$

अब हम एक और अवधारणा स्वीकार कर लेते हैं। वह यह है कि $\phi(z)$ को एक घात श्रेणी (power series) के रूप में रखा जा सकता है। यानी

$$\phi(z) = a_0 + a_1 z + a_2 z^2 + \dots \quad (8\ 17)$$

जहाँ a_0, a_1, a_2 इत्यादि ऐसे अचर (constants) हैं जो समीकरण

$$\sum_{r=1}^n \phi(z_r) = 0 \text{ को सतुष्ट कर सकें।}$$

क्योंकि $\phi(z_r) = a_0 + a_1 z_r + a_2 z_r^2 +$

$$\begin{aligned} \sum_{r=1}^n \phi(z_r) &= n a_0 + a_1 \sum_{r=1}^n z_r + a_2 \sum_{r=1}^n z_r^2 + a_3 \sum_{r=1}^n z_r^3 + \dots \\ &= 0 \end{aligned} \quad (818)$$

यह समीकरण तभी सतुष्ट हो सकता है जब उसके हर एक पद का मान शून्य हो। यदि a_1 को छोड़कर अथ a_0, a_2, a_3 इत्यादि सब शून्य हों तो भी यह सतुष्ट हो जायगा, क्योंकि

$$\sum_{r=1}^n z_r = 0$$

$$\phi(z) = \frac{df(z)/dz}{f(z)} = a_1 z$$

$$\text{अथवा } \frac{d \log f(z)}{dz} = a_1 z \quad (819)$$

परन्तु हम जानते हैं कि यदि $\log f(z) = \frac{a_1}{2} z^2 + \log C$ हो

जहाँ C कोई भी अचर है तो $\frac{d \log f(z)}{dz} = a_1 z$ हो जाता है।

इसलिए ऊपर के समीकरण में हम यह मान सकते हैं कि

$$f(z) = c e^{a_1 z^2/2}$$

आपको याद होगा कि हम यह अवधारणा लेकर चले थे कि $f(z)$ का महत्तम मान $z=0$ पर होता है और जैसे जैसे z का मान शून्य से अधिकाधिक अंतर पर होता जाता है वैसे ही वैसे $f(z)$ का मान शून्य की ओर अभसर होता जाता है। यह तभी हो सकता है जब a_1 एक ऋणात्मक संख्या हो। इसलिए हम a_1 के स्थान पर $-\frac{1}{\sigma^2}$ लिख सकते हैं—

$$\therefore f(z) = c e^{-z^2/2\sigma^2}$$

$$\text{परन्तु } \int_{-\infty}^{+\infty} f(z) dz = 1$$

$$\therefore C \int_{-\infty}^{+\infty} e^{-z^2/2\sigma^2} dz = C\sqrt{2\pi} \sigma = 1$$

$$\text{क्योंकि यह ज्ञात है कि } \int_{-\infty}^{+\infty} e^{-z^2/2\sigma^2} dz = \sigma\sqrt{2\pi}$$

$$\therefore C = \frac{1}{\sigma\sqrt{2\pi}}$$

$$\text{और } f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-z^2/2\sigma^2}$$

आप यह तो पहचान ही गये होंगे कि यह फलन एक प्रसामान्य चर का घनत्व फलन है जिसका माध्य शून्य और मानक विचलन σ है।

§ ८'७ परिकल्पनाओं की जाँच में प्रसामान्य वंटन का उपयोग

अब आप कई परिस्थितियों से परिचित हो चुके हैं जहाँ यह आसा की जा सकती है कि वटन प्रसामान्य होगा। आप यह भी समझ चुके हैं कि प्रसामान्य वटन का आविष्कार त्रुटियों के वटन के रूप में किन अवधारणाओं को लेकर हुआ था। यह कदाचित् आप समझ गये होंगे कि किसी वटन के माध्य और मानक विचलन का विशेष महत्त्व क्यों है। यदि हमें किसी यादृच्छिक चर के माध्य और मानक विचलन ज्ञात हैं और यदि हम एक काफी बड़ा प्रतिदर्श इस चर के लिए लेते हैं तो हम जानते हैं कि इस प्रतिदर्श के माध्य का वटन क्या होगा। आइए अब हम देखें कि इस वटन का उपयोग कुछ परिकल्पनाओं की जाँच के लिए किस प्रकार किया जा सकता है।

उदाहरण (१) आसाम की एक जाति में मनुष्यों की ऊँचाई का बड़े पैमाने पर अध्ययन किया गया। पता लगा कि ऊँचाई का वितरण प्रसामान्य है जिसका माध्य ५ फुट ६ इंच और मानक विचलन २.५ इंच है। कुछ इतिहासकारों का मत है कि यह जाति राजस्थान के एक विशेष भाग से लगभग दो सौ वर्ष पहले आसाम में आयी थी। यह सर्वविदित है कि इस जाति के लोग जाति के अन्दर ही विवाह करते हैं। और राजस्थान के उम भाग के लोग भी अन्य जाति या विदेशियों से विवाह नहीं करते। प्राणि-विज्ञान के ज्ञाताओं के अनुसार मनुष्य की ऊँचाई वंशानुगत गुणों पर ही अधिक निर्भर करती है। इसलिए यदि इतिहासकारों के मत में कुछ सच्चाई है तो इन दोनों जातियों के मनुष्यों की ऊँचाई का वितरण एक-ना होना चाहिए। यदि इनमें अंतर हो तो इतिहासकारों के मत से विश्वास उठ जायगा।

अब हमें इतिहासकारों के मत को एक सांख्यिकीय परिकल्पना का रूप देना होगा जिसकी जाँच की जा सके। यह सांख्यिकीय रूप निम्नलिखित हो सकता है। "राजस्थान के इस विशेष भाग की जाति में मनुष्यों की ऊँचाई का वितरण प्रसामान्य है जिसका माध्य ५ फुट ६ इंच और मानक विचलन २.५ इंच है।" इस निराकरणयोग्य परिकल्पना की जाँच के लिए इस भाग की जनसंख्या से एक यादृच्छिकीकृत प्रतिदर्श लिया गया जिसमें १०० मनुष्य थे। इन मनुष्यों की ऊँचाई नापी गयी और इस प्रतिदर्श में ऊँचाइयों के माध्य का कलन किया गया। हमने प्रसामान्य वितरण के बारे में जो कुछ अध्ययन किया है उससे हमें यह मालूम है कि $t = \left(\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \right)$ का वितरण $N(0, 1)$

है जहाँ \bar{x} प्रतिदर्श-माध्य, μ समष्टि-माध्य, σ समष्टि का मानक विचलन और n प्रतिदर्श-संख्या है। इस उदाहरण में

$$\mu = 5 \text{ फुट } 6 \text{ इंच}$$

$$\sigma = 2.5 \text{ इंच}$$

$$n = 100$$

प्रतिदर्श की ऊँचाइयों का माध्य ५ फुट ७ इंच पाया गया। अर्थात् $\bar{x} = 5 \text{ फुट } 7$ इंच और $\bar{x} - \mu = 1$ इंच

$$\therefore t = \frac{1}{2.5} \times \sqrt{100} = 4$$

$N(0, 1)$ की सारणी में देखने से हमें ज्ञात होता है कि इतना बड़ा या इससे भी बड़े मान होने की प्रायिकता 0.00005 से कम है। इस कारण हमें इस निराकरणीय* परिकल्पना को कि राजस्थान के इस भाग की जाति के मनुष्यों की ऊँचाई का वितरण प्रसामान्य है—जिसका माध्य 5 फुट 6 इंच और मानक विचलन 2.5 इंच है—स्थापने की बाध्यता होगी, परन्तु यह परिकल्पना इतिहासकारों के मत का ही निष्कर्ष है। इसलिए इसको त्यागने का अर्थ है यह समझना कि इतिहासकारों का मत गलत है।

पाठकों का ध्यान इस ओर गया होगा कि यह परिकल्पना केवल इतिहासकारों के मत पर ही निर्भर नहीं है, बल्कि प्राणिविज्ञान के ज्ञाताओं के मत से सबध रखती है। यदि उनका मत प्रमाणित नहीं हो चुका है और उसमें संदेह की कुछ गुंजाइश है तो इतिहासकार यह कह सकते हैं कि इस जाँच से यह निष्कर्ष भी निकल सकता है कि प्राणिविज्ञान का यह मत ठीक नहीं है। इस प्रकार एक ही प्रयोग के नतीजों की व्याख्या भिन्न-भिन्न लोग विभिन्न तरीकों से कर सकते हैं। ऐसी स्थिति में हमारी जाँच अर्थहीन हो जाती है। यह जाँच उसी समय कुछ अर्थ रखेगी जब जिस मत की हम पुष्टि अथवा स्पष्टन करना चाहते हैं उसके अतिरिक्त और किसी भी ऐसे मत पर निराकरणीय परिकल्पना निर्भर न करे जिसकी सच्चाई में संदेह हो।

उदाहरण (२) एक कारखाने में किमी विशेष मशीन के लिए छड़ें (rods) बनती हैं। मशीन के लिए इन छड़ों की लम्बाई १५ सेंटीमीटर होना चाहिए। इसलिए कारखाने में यहीं उद्देश्य सामने रखा जाता है। परन्तु मनुष्य, मशीन और माल के कारण कुछ-न-कुछ त्रुटि होना संभव है। अतः यह संभव नहीं है कि प्रत्येक छड़ की लम्बाई ठीक 15 सेंटीमीटर ही हो—न कम न ज्यादा। यदि इन छड़ों का निर्माण-कार्य बिल्कुल नियंत्रित है तो यह देखा जाता है कि इनकी लम्बाई का वितरण प्रसामान्य होता है जिसका माध्य 15 सेंटीमीटर और मानक विचलन 0.1 सेंटीमीटर है।

एक दिन किसी यादृच्छिक रूप से चुने हुए समय पर 16 छड़ों का एक प्रतिदर्श लिया गया। इन सबकी लम्बाई नापी गयी और उनके माध्य का कलन किया गया। यह माध्य 15.1 सेंटीमीटर था। अब तब यह करना है कि 15 सेंटीमीटर से इस माध्य का अंतर क्या यह इंगित करता है कि निर्माण-कार्य इस समय नियंत्रण से बाहर था।

*प्रयोग द्वारा जिस परिकल्पना के बारे में यह निर्णय करना होता है कि वह निराकरण करने के योग्य है अथवा नहीं उसको निराकरणीय परिकल्पना (null hypothesis) कहते हैं।

इसको तय करने के लिए पहले हम इस निराकरणिय परिकल्पना से आरभ करेगे कि निर्माणकार्य नियंत्रित था। इसका अर्थ यह होगा कि यह प्रतिदर्श एक समष्टि में से लिया गया है, जिसका वितरण प्रसामान्य $N(15, 0.1)$ है। आइए, हम देखें कि इस प्रयोग में t का मान क्या है।

$$t = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$= \frac{15.1 - 15.0}{0.1} \sqrt{16}$$

$$= 4$$

t के इतने अधिक या इससे भी अधिक मान होने की प्रायिकता हम पहले उदाहरण में ही मालूम कर चुके हैं। हम यह भी जानते हैं कि यह इतनी कम है कि निराकरणिय परिकल्पना को त्याग देना ही उचित मालूम देता है। इसलिए यह समझा जा सकता है कि निर्माण वास्तव में नियंत्रण से बाहर था।

उदाहरण (३) मनुष्यों की बुद्धि को नापने के लिए एक प्रकार का परीक्षण तैयार किया गया है जिसे बुद्धि-परीक्षण (intelligence test) कहते हैं। इसमें 200 या 300 छोटे छोटे प्रश्न पूछे जाते हैं जिनके उत्तर एक निश्चित समय में देने होते हैं। इन उत्तरों पर नम्बर दिये जाते हैं और यदि किसी को इस परीक्षा में 60 प्रतिशत से कम नम्बर मिले तो उसे असतोपजनक समझा जाता है। एक विश्वविद्यालय की ओर से 20 वर्ष पूर्व इस परीक्षा का उपयोग हजारों विद्यार्थियों पर किया गया था। यह देखा गया कि इस प्रतिशत विद्यार्थियों का परीक्षा-फल असतोपजनक था। इस वर्ष इस परीक्षा का उपयोग 64 विद्यार्थियों पर किया गया था। इनमें से केवल पाँच ऐसे थे जिनका परीक्षाफल असतोपजनक था।

एक वैज्ञानिक या कहना है कि इस प्रयोग से यह मालूम होता है कि कुल मनुष्यों में बुद्धिमान् मनुष्यों का अनुपात जितना 20 वर्ष पूर्व था उससे आज अधिक है। यहाँ बुद्धिमान् मनुष्यों से वैज्ञानिकों का तात्पर्य उन मनुष्यों में है जिन्हें बुद्धि परीक्षा में 60 प्रतिशत से अधिक नम्बर मिले। हमें यह देखना है कि इस वैज्ञानिक का कथन कहा तक युक्तियुक्त है।

पाठक निश्चय ही यह सोचेंगे कि ऐसी स्थिति में द्विपद-वटन का उपयोग करना चाहिए, क्योंकि हमें यह जाँच करनी है कि इस प्रतिदर्श में बुद्धिमान् मनुष्यों का जो अनुपात है उतना या उससे अधिक अनुपात होने की प्रायिकता क्या है। यदि यह समझ

लिया जाय कि अब भी समष्टि में अनुपात 90 प्रतिशत ही हैं तो पाठक का यह विचार ठीक है। परन्तु द्विपद-वटन में प्रयोग में कुछ कठिनाई है। जैसा कि पहले लिखा जा चुका है N का 50 से अधिक मान के लिए द्विपद-वटन की कोई सारणी प्रस्तुत नहीं है। इसलिए द्विपद-वटन के प्रयोग के लिए स्वयं इस प्रायिकता का कलन करना होगा। यद्यपि यह कठिन नहीं है परन्तु इसमें बहुत समय लगेगा। इस कारण द्विपद वटन के स्थान में हम इस वटन के किसी सन्निकटन (approximation) का उपयोग कर सकते हैं जिससे ऊपर दी हुई निराकरणाय परिकल्पना की जाँच कुछ मिनटों में ही हो सकती है।

द्विपद-वटन का माध्य है

$$Np = 64 \times 0.10 \\ = 6.4$$

$$\text{इसका मानक विचलन है } \sqrt{Npq} = \sqrt{64 \times 0.10 \times 0.90} \\ = 8 \times 0.30 \\ = 2.40$$

इसलिए इस द्विपद-वटन का सन्निकटन एक प्रसामान्य वटन से किया जा सकता है जिसका माध्य 6.4 और मानक विचलन 2.4 है। क्योंकि विद्यार्थियों के इस प्रतिशत में अमतोपजनक फल पानेवाला की संख्या x का यह वटन है

$$\text{इसलिए } f = \frac{n - 6.4}{2.4} \text{ का वटन प्रसामान्य है जिसका माध्य 0 और मानक}$$

विचलन 1 है।

$$f \text{ का प्रेक्षित मान है } = \frac{5 - 6.4}{2.4} \\ = -\frac{1.4}{2.4} \\ = -0.583$$

f का इतना कम या इससे भी कम मान के होने की प्रायिकता 30% से भी अधिक है। इसलिए यदि कम-बुद्धिमान् मनुष्या की प्रतिशतता अब भी 10% ही हो, फिर भी हम सो बार में तीस बार यह उम्मीद कर सकते हैं कि 64 विद्यार्थियों के प्रति-दर्श में 5 या उससे भी थोड़े कम-बुद्धिमान् विद्यार्थी पाये जायेंगे। यह प्रायिकता इतनी

अधिक है कि इस प्रयोग से इतना बड़ानिष्कर्ष निकाल लेना युक्तियुक्त मालूम नहीं होता कि अब दुद्धिमान् मनुष्यों का अनुपात बढ़ गया है।

यद्यपि प्रसामान्य वटन के अनेको और विभिन्न उपयोग हैं, परन्तु आप अब तक परिकल्पना की जाँच में इसके उपयोग को काफी समझ चुके होंगे। और अधिक उदाहरण देने की आवश्यकता नहीं है, क्योंकि चाहे किसी विज्ञान में या किसी परिकल्पना की जाँच के लिए इसका प्रयोग किया जाय सिद्धान्त और तरीका वही रहेगा।

परन्तु यदि आपका दृष्टिकोण आलोचनात्मक है तो आपको प्रसामान्य वटन और प्वासो वटन के उपयोग के बारे में एक सदेह अवश्य उठा होगा। इन उपयोगों में आपका ध्यान इस ओर गया होगा कि कई बार मूल समस्या यह नहीं होती कि प्रतिदर्श एक विशेष प्रसामान्य अथवा प्वासो समष्टि में लिया गया है। बल्कि वह केवल समष्टि के माध्य अथवा मानक विचलन में सबब रखती है। प्रायः सभी उदाहरणों में हमने यह कहा है कि एक बहुत बड़े प्रतिदर्श के आधार पर हम यह जानते हैं कि वटन प्वासो है अथवा प्रसामान्य है या वह आयताकार है। लेकिन यह स्पष्ट है कि इस बड़े प्रतिदर्श में चर का वटन ठीक प्रसामान्य अथवा प्वासो होना असंभव है। इस प्रतिदर्श में चर के वास्तविक वटन और गणितीय वटन में अन्तर के महत्त्व को मापने के लिए भी तो कोई परीक्षण होना चाहिए। इसका विवरण हम अगले अध्याय में देंगे जिसमें हमारा परिचय एक नये वटन χ^2 -वटन (काई-बर्ग वटन) से होगा। जिस समस्या का यहाँ हमने उल्लेख किया है उसके अलावा अन्य समस्याओं के सुलझाने में उसके प्रयोग का वर्णन भी वहाँ किया जायेगा।

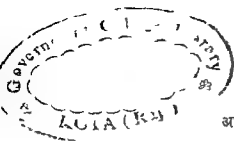
सारणी संख्या 82

प्रसामान्य वटन $N(\mu, \sigma)$ के कुछ प्रतिशतता बिंदु

प्रतिशतता	50	25	10	05
$\frac{x-\mu}{\sigma}$	1.65	1.96	2.33	2.58

विस्तृत सारणी के लिए देखिए

“Statistical Tables for Biological, Agricultural and Medical Research” By Fisher and Yates



अध्याय ९

X^2 -घंटन

§ ९.१ यादृच्छिक चर के फलन का घंटन

मान लीजिए कि एक यादृच्छिक चर X का घनत्व-फलन $f(x)$ है। यदि $g(X)$ इस चर का कोई एकत्वनी* (monotonic) फलन हो तो इस फलन का घनत्व-फलन क्या होगा? यदि हम इसको $f_1(x)$ से सूचित करें तो

$$\begin{aligned} f_1(x) &= \lim_{\epsilon \rightarrow 0} \frac{P[x < g(X) < x + \epsilon]}{\epsilon} \\ &= \lim_{\epsilon \rightarrow 0} \frac{P[g^{-1}(x) < X < g^{-1}(x + \epsilon)]}{\epsilon} \end{aligned}$$

यहाँ $g^{-1}(x)$ से हम X के उस मान को सूचित करते हैं जिसके लिए $g(X) = x$ हो। क्योंकि हमें X का घनत्व-फलन ज्ञात है, इसलिए

$$P[g^{-1}(x) < X < g^{-1}(x + \epsilon)]$$

का परिवर्तन किया जा सकता है।

(देखिए § ४.२१)

§ ९.२ X^2 का घंटन

ऊपर दिये साधारण नियम का एक बहुत ही सरल उदाहरण वह है जब

$$\begin{aligned} g(X) &= X^2 \\ g^{-1}(x) &= +\sqrt{x} \text{ और } -\sqrt{x} \end{aligned}$$

$$\text{परन्तु } \sqrt{x + \epsilon} = (x + \epsilon)^{1/2} = x^{1/2} \left[1 + \frac{\epsilon}{2x} + \frac{1}{2} \left(\frac{\epsilon}{2x} \right)^2 + \dots \right]$$

*यदि x का कोई फलन $g(x)$ ऐसा हो जिसका मान x के बढ़ने के साथ बिना घटे बढ़ता जाय अथवा बिना बढ़े घटता जाय तो उस फलन को x का एकत्वनी फलन कहते हैं।

यदि ϵ बहुत छोटा हो तो ϵ^2 और ϵ के अन्य ऊँचे घातों (powers) की उपेक्षा की जा सकती है।

$$\therefore \sqrt{x + \epsilon} = x^{1/2} \left(1 + \frac{\epsilon}{x} \right)$$

$$\begin{aligned} \therefore P[\lambda < X^2 < x + \epsilon] &= P \left[-x^{1/2} - \frac{1}{2} \epsilon x^{-1/2} < X < -x^{1/2} \right] \\ &\quad + P \left[x^{1/2} < X < x^{1/2} + \frac{1}{2} \epsilon x^{-1/2} \right] \\ &= \frac{1}{2} \epsilon x^{-1/2} \left[f(x^{1/2}) + f(-x^{1/2}) \right] \quad (9.1) \end{aligned}$$

यदि X का वटन $N(0,1)$ हो तो

$$\begin{aligned} P[\lambda < X^2 < \lambda + \epsilon] &= \frac{1}{2} \epsilon x^{-1/2} \left[\frac{1}{\sqrt{2\pi}} e^{-x/2} + \frac{1}{\sqrt{2\pi}} e^{-x/2} \right] \\ &= \frac{1}{\sqrt{2\pi}} \epsilon x^{-1/2} e^{-x/2} \end{aligned}$$

• यदि X का वटन $N(0,1)$ हो तो X^2 का घनत्व फलन

$$f_1(X) = \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-x/2} \quad (9.2)$$

यह वटन 1 स्वातन्त्र्य-संख्या (degree of freedom 1) वाला χ^2 -वटन कहलाता है। जिस चर का ऐसा वटन होता है उसे χ_1^2 चर कहते हैं।

§ 9.3 χ_n^2 चर की परिभाषा

इस प्रकार के n स्वतंत्र χ_1^2 चरों के योग को χ_n^2 से सूचित करते हैं और इस चर को n स्वातन्त्र्य-संख्या वाला χ^2 -चर कहा जाता है। यह सिद्ध किया जा सकता है कि इस चर का घनत्व-फलन $f_n(x)$ निम्नलिखित होता है।

$$f_n(x) = \frac{1}{2^{n/2}} \Gamma\left(\frac{n}{2}\right) x^{n/2-1} e^{-x/2} \quad (9.3)$$

यहाँ $\Gamma\left(\frac{n}{2}\right)$ निम्नलिखित समाकल (integral) का मान है

$$\Gamma\left(\frac{n}{2}\right) = \int_0^{\infty} e^{-x} x^{\frac{n}{2}-1} dx$$

यह स्पष्ट है कि X_n^2 केवल धनात्मक मान ही धारण कर सकता है और सब धनात्मक मानों को धारण कर सकता है। क्योंकि $f_n(x)$ हम यादृच्छिक चर का घनत्व-फलन है इसलिए

$$\int_0^{\infty} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-x/2} dx = 1$$

$$\text{यानी } \int_0^{\infty} x^{\frac{n}{2}-1} e^{-x/2} dx = 2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right) \quad (94)$$

यह फल n के प्रत्येक मान के लिए सत्य है।

§ ९४ X^2 वटन के कुछ गुण

यदि X का वटन X_n^2 है तो

$$\begin{aligned} E(X) &= \int_0^{\infty} x \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-x/2} dx \\ &= \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \times 2^{\frac{n+2}{2}} \Gamma\left(\frac{n+2}{2}\right) \\ &= 2 \frac{\Gamma\left(\frac{n+2}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \end{aligned}$$

$\Gamma(x)$ एक फलन है जिसमें कुछ विशेषताएँ हैं। उनमें से एक यह है कि $\Gamma(x+1) = x \Gamma(x)$ । यह x के सब घनात्मक मानों के लिए सत्य है। इसलिए

$$\Gamma\left(\frac{n+2}{2}\right) = \frac{n}{2} \Gamma\left(\frac{n}{2}\right)$$

$$\therefore E(x) = n \quad (9.5)$$

इस प्रकार हम देखते हैं कि किसी χ^2 -बटन का माध्य उसकी स्वातन्त्र्य-संख्या के बराबर होता है।

$$(2) \quad V(X) = E(X^2) - E^2(X)$$

$$E(X^2) = \int_0^{\infty} x^2 \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx$$

$$= \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \int_0^{\infty} x^{\frac{n+4}{2}-1} e^{-\frac{x}{2}} dx$$

$$= \frac{2^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \frac{n+2}{2} \frac{n}{2} \Gamma(\frac{n}{2})$$

$$= n(n+2)$$

$$\therefore V(X) = n(n+2) - n^2$$

$$= 2n$$

$$(9.6)$$

(3) दो स्वतंत्र χ^2 -चरों का योग

मान लीजिए कि X_1 कोई $\chi^2_{n_1}$ -चर है और X_2 कोई $\chi^2_{n_2}$ -चर है और ये दोनों चर एक दूसरे से स्वतंत्र हैं। इनको क्रमशः n_1 तथा n_2 चरों का योग समझा जा सकता है जो एक दूसरे से स्वतंत्र हो और जिनका बटन χ^2 की जाति का हो। इसलिए

इन दो चरों का योग $(X_1 + X_2)$ एक $\chi^2_{n_1+n_2}$ चर है।

इसी प्रकार कई स्वतंत्र χ^2 -चरों का योग भी χ^2 -चर होता है और उसकी स्वातन्त्र्य-संख्या इन विभिन्न χ^2 -चरों की स्वातन्त्र्य-संख्याओं के योग के बराबर होती है।

§ ९.५ समष्टि को पूर्ण रूप से विनिर्दिष्ट (*specify*) करनेवाली परिकल्पनाओं के लिए χ^2 परीक्षण

यदि निराकरणयोग्य परिकल्पना समष्टि को पूर्ण रूप से विनिर्दिष्ट करती हो और यदि इस समष्टि से चुना हुआ एक यथेष्ट परिमाण का यादृच्छिक प्रतिदर्श आप के पास हो तो इस परिकल्पना की जाँच आप कैसे करेंगे? मान लीजिए कि परिकल्पना यह है कि समष्टि $N(\mu, \sigma)$ है। इसके लिए एक परीक्षण का परिचय आप प्रसामान्य वटन के उपयोग के सबंध में पा चुके हैं। परंतु वह परीक्षण किसी हद तक समष्टि के माध्य μ से अधिक सबंध रखता था। यदि प्रतिदर्श का माध्य μ के बराबर अथवा उसके अत्यंत निकट होता तो समष्टि के प्रसामान्य न होंते हुए भी हम उस परीक्षण द्वारा परिकल्पना के विरुद्ध फंसला नहीं दे सकते थे। यदि समष्टि प्रसामान्य भी होनी परंतु उसका वास्तविक प्रसरण परिकल्पित प्रसरण σ^2 से बहुत अधिक होता तो भी वह परीक्षण इसकी जाँच नहीं कर सकता था। निश्चय ही आप ऐसे परीक्षण की खोज में होंगे जिसका सबंध पूरे वटन से हो न कि केवल उसके माध्य से। ऐसा एक परीक्षण सांख्यिकी ने खोज निकाला है। यह न केवल प्रसामान्य अथवा प्वासो वटनों से संबंधित है वरन् प्रायः किसी भी वटन से संबंधित परिकल्पना की जाँच के लिए उपयुक्त है। इस परीक्षण में χ^2 वटन का प्रयोग किया जाता है और इसके लिए काफी बड़े प्रतिदर्शों की आवश्यकता होती है।

मान लीजिए कि यादृच्छिक चर जितने मान धारण कर सकता है उन सबके कुलक (set) को S से सूचित किया जाता है। मान लीजिए इस कुलक को r भागों में विभाजित कर दिया जाता है, जिनको क्रमशः S_1, S_2, \dots, S_r से सूचित किया जायगा। उदाहरण के लिए यदि यादृच्छिक चर का वटन द्विपद है जिसके प्राचल (parameters) 6 और p हैं तो S निम्नलिखित मानोवाला कुलक है—

$$0, 1, 2, 3, 4, 5 \text{ और } 6$$

यही वे मान हैं जो कि ऊपर दिया हुआ द्विपद चर धारण कर सकता है। इन सात मानों के कुलक को सुविधानुसार कई भागों में विभाजित किया जा सकता है। यथा, मान लीजिए पहिले भाग में $0, 1$ और 2 हैं, दूसरे में 3 , तीसरे में 4 , और चौथे में 5 तथा 6 । ये भाग परस्पर अपवर्जी (mutually exclusive) तथा निःशेषी (exhaustive) हैं अर्थात् S का प्रत्येक मान किसी-न-किसी भाग में सम्मिलित हो गया है।

हम यादृच्छिक चर के वटन के आधार पर उसके इन विभिन्न भागों में होने की प्रायिकता का परिकलन कर सकते हैं। ये प्रायिकताएँ निम्नलिखित हैं

$$P(S_1) = (1-p)^6 + 6(1-p)^5 p + 15(1-p)^4 p^2$$

$$P(S_2) = 20(1-p)^3 p^3$$

$$P(S_3) = 15(1-p)^2 p^4$$

$$P(S_4) = 6(1-p)p^5 + p^6$$

यदि हम $P(S_i)$ को p_i द्वारा सूचित करें तो

$$\sum_{i=1}^4 p_i = 1$$

क्योंकि ये कुलक परस्पर अपवर्जी तथा निःशेषी हैं।

मान लीजिए कि n परिमाण का एक यादृच्छिक प्रतिदर्श चुना जाता है और इन विभिन्न कुलकों में चर के प्रेक्षित मानों की मख्या क्रमशः y_1, y_2, \dots, y_r है।

हमारा पहिला उद्देश्य तो एक ऐसे माप को मालूम करना है जो प्रतिदर्श-वटन तथा परिकल्पित वटन के अंतर का आभास दे सके। परिकल्पना के आधार पर प्रतिदर्श में प्रत्याशित बारबारता क्रमशः

$$np_1, np_2, \dots, np_r$$

थी। जो माप हम चाहते हैं उसे स्पष्टतया इन प्रत्याशित बारबारताओं और प्रेक्षित बारबारताओं के अंतरों का फलन होना चाहिए। इस प्रकार का एक फलन निम्नलिखित है

$$\begin{aligned} \chi^2 &= \sum_{i=1}^r \frac{(y_i - np_i)^2}{np_i} \\ &= \sum_{i=1}^r \frac{y_i^2}{np_i} - n \quad \dots \dots (97) \end{aligned}$$

कार्ल पियरसन (Karl Pearson) ने यह सिद्ध किया था कि इस ऊपर लिखित माप को कुछ विशेषताएँ हैं। जैसे-जैसे प्रतिदर्श परिमाण n को बढ़ाया जाय इस माप का वटन ऐसे χ^2 -वटन की ओर अग्रसर होता जाता है जिसकी स्वातन्त्र्य-सख्या $(r-1)$ है। इस वटन की उपपत्ति (proof) यहाँ नहीं दी जा रही है।

इस गुण के प्रयोग से एक और परिकल्पना-परीक्षण तैयार कर सकते हैं जिससे इस परिकल्पना का परीक्षण किया जा सकता है कि यादृच्छिक चर के विभिन्न कुलकों

में होने की प्रायिकताएँ क्रमशः p_1, p_2, \dots, p_r हैं। यह निराकरणयोग्य परिकल्पना स्वयं एक विदीप वटन पर आधारित है। यदि इस निराकरणयोग्य परिकल्पना को सदेह-जनक समझा जाता है तो इस आधार वटन पर सदेह होना भी स्वाभाविक है।

मान लीजिए $\chi^2_{r-1}(p)$ द्वारा हम उम मान को सूचित करते हैं जिससे अधिक होने की प्रायिकता—किन्हीं χ^2_{r-1} चर के लिए— p प्रतिशत है। यदि p इतना छोटा हो कि इतनी कम प्रायिकता वाली घटना का होना प्रायः अमभव समझा जाय और यदि प्रतिदर्श-परिमाण इतना अधिक हो कि χ^2 को एक χ^2_{r-1} चर माना जा सके तो हम आगा करते हैं कि यदि परिकल्पना सत्य है तो χ^2 का मान $\chi^2_{r-1}(p)$ से अधिक नहीं होगा। यदि χ^2 का प्रेक्षित मान $\chi^2_{r-1}(p)$ से अधिक हो तो हम परिकल्पना पर मदेह करने और उसको त्यागने के लिए बाध्य हो जाते हैं। इस सख्या p को इस परीक्षण का सार्थकता-स्तर (level of significance) कहते हैं।

§ ९.६ χ^2 वटनों की सारणी

अनुभव से ज्ञात हुआ है कि यदि प्रतिदर्श-परिमाण इतना अधिक हो कि प्रत्येक प्रत्याशित आवृत्ति np , पाँच या पाँच से अधिक हो तो हम χ^2 -वटन का प्रयोग कर सकते हैं। यदि किसी कुलक में प्रत्याशित बारंबारता पाँच से कम होती है तो उस कुलक को समीप के किमी अन्य कुलक से मिला दिया जाता है जिसमें इन बड़े हुए कुलक में प्रत्याशित बारंबारता पाँच या उससे अधिक हो जाय। सांख्यिकी ने इस प्रकार के परीक्षण के लिए एक सारणी बना रखी है। इसमें 1 से 30 तक की स्वानुक्रम-सख्याओं वाले χ^2 वटनों के लिए, तथा p के विभिन्न मानों के लिए $\chi^2(p)$ के मान दिये हुए हैं। इस सारणी का उपयोग केवल उसी स्थिति में किया जाता है जब स्वातन्त्र्य-सख्या $(r-1)$ तीस या तीस से कम हो। यदि यह तीस से भी अधिक हो तो हम रोनाल्ड ए फिसर द्वारा खोजे हुए इस गुण का प्रयोग कर सकते हैं कि n के बड़े मानों के लिए $\sqrt{2\chi^2/n}$ का वटन प्रायः प्रसामान्य होता है और उमका माध्य $\sqrt{2n-1}$ तथा प्रसरण 1 होता है।

§ ९७ आइए अब हम दो-तीन उदाहरणों द्वारा इस सिद्धान्त को अच्छी तरह से समझ लें

उदाहरण (१) कुछ लोग का विश्वास है कि विभिन्न ग्रह और अन्य आकाशमण्डल सप्ताह के अलग-अलग दिनों पर राज्य करते हैं। वे ये भी विश्वास करते हैं कि इन ग्रहों का वर्षों पर अलग अलग प्रभाव पड़ता है। इस तरह वे भाशा करते हैं कि यदि कुल वर्षों के दिनों की जांच की जाय तो मालूम होगा कि उनमें सोमवार की अपेक्षा इतवार अधिक है, मंगलवार की अपेक्षा सोमवार अधिक है इत्यादि। यानी विभिन्न वारों की बारबारताएँ भिन्न भिन्न होंगी। हम यहाँ उपयुक्त मूल विश्वास की विवेचना नहीं करना चाहते बल्कि उस विश्वास से सवधित वर्षों के दिनों के बारे में एक सांख्यिकीय परिकल्पना की जाँच से ही सतोप कर लेंगे।

हमारी निराकरणीय परिकल्पना H_0 यह है कि वर्षों की रविवार, सोमवार, मंगलवार, बुधवार, वृहस्पतिवार, शुकवार एवं शनिवार को होने की प्रायिकताएँ समान हैं। यदि इन प्रायिकताओं को $p_1, p_2, p_3, p_4, p_5, p_6$ और p_7 से सूचित किया जाय तो H_0 यह है कि $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = \frac{1}{7}$ और इसका निष्कर्ष यह है कि यदि हम पिछले कई वर्षों के दिनों के आँकड़ों का विश्लेषण करें तो उनमें सप्ताह के प्रत्येक वार का प्रतिनिधित्व लगभग समान होगा।

प्रयोग—किसी विशेष स्थान के मौसम वैज्ञानिक पत्र (meteorological office) से हम पिछले 301 वर्षों के दिनों का विश्लेषण करके उनमें विभिन्न वारों की बारबारता का पता लगायेंगे।

सांख्यिकता-स्तर (level of significance)

हम यह पहिले से ही तय कर लेते हैं कि यदि प्रेक्षित बारबारताओं की इस परिकल्पना के आधार पर परिकल्पित प्रायिकता पाँच प्रतिशत से कम होगी तो हम परिकल्पना का त्याग कर देंगे। इसलिए इस प्रयोग का सांख्यिकता-स्तर $p = 5$ प्रतिशत है।

अस्वीकृति-क्षेत्र (region of rejection)

यदि χ^2 -का प्रेक्षित मान χ_0^2 के सारणी में दिये हुए पांच प्रतिशत बिंदु 12.592 से अधिक हो तो हम निराकरणीय परिकल्पना H_0 को त्याग देंगे अथवा उसे अस्वीकार करेंगे।

(देखिए मारपी सख्या 98)

आंकड़े (data) —

वर्षों के दिनों की सात कुलका में विभाजित किया गया है। हर एक कुलक सप्ताह के एक विशेष वार की हुई वर्षों से संबंधित है। नीचे सारणी में इन कुलका में प्रेक्षित वारवारताएँ दी हुई हैं। निराकरण योग्य परिकल्पना के अनुसार हर एक कुलक की प्रत्याक्षित वारवारता $\frac{301}{7} = 43$ है।

सारणी सख्या 9 1

पिछले 301 वर्षों के दिनों में विभिन्न वारों की वारवारता

रविवार	सोमवार	मंगलवार	बुधवार	बृहस्पतिवार	शुक्रवार	शनिवार
(1)	(2)	(3)	(4)	(5)	(6)	(7)
55	43	37	48	52	34	32

विश्लेषण —

$$\begin{aligned}
 \chi^2 &= \sum_{i=1}^7 \frac{(v_i - np_i)^2}{np_i} && \text{(देखिए समीकरण 9 7)} \\
 &= \frac{1}{43} [(55)^2 + (43)^2 + (37)^2 + (48)^2 + (52)^2 + (34)^2 + (32)^2] \\
 &= \frac{1}{43} [3025 + 1849 + 1369 + 2304 + 2704 + 1156 + 1024] \\
 &= \frac{14431}{43} \\
 &= 335.6
 \end{aligned}$$

फल—क्योंकि χ^2 का प्रेक्षित मान 12.592 से कम है इसलिए इन आँकड़ों के आधार पर निराकरण योग्य परिकल्पना को अस्वीकार करने का कोई कारण नहीं है।

उदाहरण (२)

अब हम फिर उस उदाहरण को लेते हैं जिसमें हमने इतिहासकारों के मत का प्रामाण्य बटन द्वारा परीक्षण किया था। इसमें निराकरण योग्य परिकल्पना यह थी कि राजस्थान के एक विशेष भाग के लोगों की ऊँचाई का बटन प्रामाण्य है जिसका माध्य 5 फुट 6 इंच और मानक-विचलन 2.5 इंच है।

हम पहिले ऊँचाई h के परास (range) को आठ भागों में विभाजित करते हैं

- (1) $h < 4$ फुट 10.5 इंच
- (2) 4 फुट 10.5 इंच $\leq h < 5$ फुट 1 इंच
- (3) 5 फुट 1 इंच $\leq h < 5$ फुट 3.5 इंच

- (4) 5 फुट 3.5 इंच $\leq h <$ 5 फुट 6 इंच
 (5) 5 फुट 6 इंच $\leq h <$ 5 फुट 8.5 इंच
 (6) 5 फुट 8.5 इंच $\leq h <$ 5 फुट 11 इंच
 (7) 5 फुट 11 इंच $\leq h <$ 6 फुट 1.5 इंच
 (8) $h \geq$ 6 फुट 1.5 इंच

नीचे की सारणी में राजस्थान के उस भाग के एक 200 परिमाण के यादृच्छिक प्रतिदर्श में इन आठ भागों के लिए बारबारताएँ दी हुई हैं। इन प्रेक्षित बारबारताओं के नीचे प्रत्याशित बारबारताएँ भी दी हुई हैं जिनका परिकलन निराकरणीय परिकल्पना के आधार पर किया गया है।

सारणी संख्या 9.2

भाग	1	2	3	4	5	6	7	8
प्रेक्षित बारबारता	3	16	23	60	65	18	14	1
प्रत्याशित बारबारता	0.27	4.28	27.18	68.27	68.27	18	4.28	0.27

अस्वीकृति-क्षेत्र — हम ऊपर के आँकड़ों के विरलेपण से पहिले ही यह तय कर चुके हैं कि यदि प्रेक्षित \bar{x}^2 का मान समुचित \bar{x}^2 -वटन के पाँच-प्रतिशत-बिंदु से अधिक होगा तो निराकरणीय परिकल्पना को अस्वीकार कर दिया जायगा।

हम यह देखते हैं कि पहिले दो और अंतिम दो कुलकों में प्रत्याशित बारबारताएँ पाँच से कम हैं। इसलिए \bar{x}^2 का परिकलन करने से पूर्व पहिले, दूसरे और तीसरे कुलकों को मिलाकर तथा छठवें, सातवें और आठवें कुलकों को मिलाकर इतने बड़े कुलक बना लेना चाहिए कि प्रत्याशित बारबारता पाँच में अधिक हो जाय। इस प्रकार कुल चार कुलक रह गये और यदि प्रेक्षित \bar{x}^2 का मान \bar{x}_3^2 के पाँच-प्रतिशत-बिंदु 7.815 से अधिक हो तो हम निराकरणीय परिकल्पना को अस्वीकार करेंगे। (देखिए सारणी संख्या 9.8)

$$\text{विरलेपण} - \bar{x}^2 = \frac{(42.00 - 31.73)^2}{31.73} + \frac{(60.00 - 68.27)^2}{68.27}$$

$$\begin{aligned}
 & + \frac{(65\ 00 - 68\ 27)^2}{68\ 27} + \frac{(33\ 00 - 31\ 73)^2}{31\ 73} \\
 & = \frac{105\ 47 + 5\ 15}{31\ 73} + \frac{68\ 39 + 10\ 69}{68\ 27} \\
 & < 7\ 815
 \end{aligned}$$

निष्कर्ष—नयोक χ^2 का प्रेक्षित मान 7 815 से कम है इसलिए इन आँकड़ों के आधार पर परिकल्पना अस्वीकृत करने का कोई कारण नहीं है।

§ ९८ आसजन-सौष्ठव का χ^2 -परीक्षण

आपका ध्यान सभवत एक बात पर गया हो कि ऊपर की निराकरणीय परिकल्पना को बिना किसी परीक्षण के ही अस्वीकृत किया जा सकता था। किसी भी प्रसामान्य वटन में ऋणात्मक मान धारण करने की प्रायिकता शून्य नहीं होगी जब कि ऊँचाई के लिए यह प्रायिकता अवश्य ही शून्य है। ऋणात्मक ऊँचाई अर्थ-शून्य है। वास्तव में जब हम यह कहते हैं कि ऊँचाई का वटन प्रसामान्य है तो इसका तात्पर्य केवल यह होता है कि वटन प्रसामान्य वटन से इतना अधिक सादृश्य रखता है कि किसी भी अथ-पूर्ण परास में ऊँचाई की बारबारता का कलन प्रसामान्य वटन के आधार पर करने से कोई विशेष त्रुटि नहीं होगी। प्रायः जब हम समष्टि के एक विशेष गणितीय वटन होने की परिकल्पना का परीक्षण करते हैं इसी प्रकार के तर्क का प्रयोग किया जाता है। कोई भी सांख्यिक कर्मी भी गभीरता से यह विचार नहीं कर सकता कि यह परिकल्पना एकदम यथार्थ हो सकती है। इस परीक्षण का तात्पर्य केवल यह जानना है कि यह विशेष गणितीय वटन समष्टि का अच्छा खासा सतौपजनक विवरण दे सकता है अथवा नहीं।

इस प्रकार के परीक्षण को आसजन-सौष्ठव (goodness of fit) का χ^2 -परीक्षण कहते हैं।

§ ९९ समष्टि को अपूर्ण रूप से विनिर्दिष्ट करनेवाली परिकल्पनाओं के लिए χ^2 -परीक्षण

ऊपर के उदाहरण में परिकल्पना में समष्टिक के μ और σ के मानों के द्वारा समष्टि को पूर्ण-रूप से विनिर्दिष्ट किया हुआ था। कुछ परिकल्पनाएँ इतनी स्पष्ट

नहीं होती। वे यह नहीं बताती कि समष्टि क्या है वरन् केवल उसके रूप (shape) से सबंध रखती हैं। उदाहरण के लिए हमारी परिकल्पना यह हो सकती है कि ऊँचाइयो का वटन प्रसामान्य है। उसके माध्य और प्रसरण को हम विनिर्दिष्ट नहीं करते। इस परिकल्पना का परीक्षण X^2 -वटन की सहायता से किस प्रकार किया जाता है, यह नीचे के उदाहरण में दिया हुआ है।

निराकरणोप परिकल्पना H_0 : राजस्थान के एक विशेष भाग के निवासियों की ऊँचाइयो का वटन प्रसामान्य है।

पूर्व इसके कि हम X^2 -परीक्षण का प्रयोग करें, हमें यह मालूम करना है कि कौन सा प्रसामान्य वटन प्रतिदर्श वटन से अधिकतम सादृश्य रखता है। इसके लिए सर्व-प्रथम हमें प्रतिदर्श-वटन से μ और σ का प्राक्कलन करना है। फिर हम इन प्राक्कलित μ और σ वाले प्रसामान्य वटन के लिए X^2 -परीक्षण करेंगे।

इसमें X^2 की स्वातन्त्र्य-संख्या कुल कुलको से एक नहीं बल्कि दो कम होती है। स्वातन्त्र्य-संख्या के मालूम करने का साधारण नियम यह है कि कुल कुलको की संख्या में से उन प्राक्कलो की संख्या को घटा दिया जाय जिनका प्राक्कलन प्रतिदर्श पर ही आधारित हो।

आँकड़े—प्रतिदर्श में माध्य 5 फुट 7 इंच और मानक-विक्षलन 2.3 इंच है। पिछले उदाहरण की भाँति ऊँचाइयो के परास को चार भागों में विभाजित किया हुआ है।

- (1) $h < 5$ फुट 4.7 इंच
- (2) 5 फुट 4.7 इंच $< h < 5$ फुट 7 इंच
- (3) 5 फुट 7 इंच $< h < 5$ फुट 9.3 इंच
- (4) $h \geq 5$ फुट 9.3 इंच

इन चार भागों में प्रेक्षित और प्रत्याशित बारबारताएँ नीचे की सारणी में दी हुई हैं।

सारणी संख्या 93

ऊँचाई कुलक	1	2	3	4
प्रेक्षित बारबारता	41	63	69	27
प्रत्याशित बारबारता	31.73	68.27	68.27	31.73

अस्योक्तित क्षेत्र—यदि प्रेक्षित X^2 का मान X_2^2 के पाँच प्रतिशत बिंदु 5.991 से

अधिक होगा तो निराकरणोप परिकल्पना को अस्वीकार कर दिया जायगा।
(देखिए मारणी सख्या ७४)

$$\begin{aligned} \text{विश्लेषण— } \chi^2 &= \frac{(41\ 00 - 31\ 73)^2}{31\ 73} + \frac{(63\ 00 - 68\ 27)^2}{68\ 27} \\ &+ \frac{(69\ 00 - 68\ 27)^2}{68\ 27} + \frac{(27\ 00 - 31\ 73)^2}{31\ 73} \\ &= \frac{85\ 93 + 22\ 37}{31\ 73} + \frac{27\ 77 + 0\ 53}{68\ 27} \\ &< 5\ 991 \end{aligned}$$

इसलिए इस परीक्षण के आधार पर परिकल्पना को अस्वीकार करने का कोई कारण नहीं है।

§ ९.१० गुण-साहचर्य (*Association of attributes*) के लिए दो स्वतंत्र प्रतिदर्शों χ^2 -परीक्षण

अब हम एक बहुत ही मनोरंजक प्रहेलिका या हल ढूँढ़ेंगे। कुछ गुण ऐसे होते हैं जिनमें परस्पर साहचर्य (association) होता है। इसका अर्थ यह है कि यदि किसी इकाई में इनमें से एक गुण विद्यमान हो तो उसमें दूसरे गुण के होने की संभावना उच्च अन्य इकाई की अपेक्षा अधिक होती है जिसमें यह पहला गुण विद्यमान न हो।

गुण-साहचर्य का एक महत्वपूर्ण उदाहरण टीके (inoculation) के प्रभाव पर विचार करने से मिलता है।

सब मनुष्यों को दो भागों में विभाजित किया जा सकता है—(१) वे जिनके टीका लग चुका हो, (२) वे जिनके टीका न लगा हो।

इन सब मनुष्यों को एक दूसरी रीति से भी दो कुलों में बाँटा जा सकता है। (१) वे जिन्हें एक निश्चित समय के अन्दर बीमारी हुई हो, (२) वे जिन्हें उसी समय में बीमारी न हुई हो।

डाक्टरों का कहना यह है कि टीका लगाने से बीमारी से बचाव होता है। उनके इस कथन की जाँच करने के लिए सांख्यिक दो यादृच्छिक प्रतिदर्श ले सकता है—एक उन मनुष्यों में से जिनके टीका लग चुका हो और दूसरा उन मनुष्यों में से जिनके टीका न लगा हो। यदि टीके का कुछ भी प्रभाव बीमारी को रोकने पर नहीं पड़ता तो इन दोनों प्रतिदर्शों में बीमारों का प्रत्याशित अनुपात समान होगा। यदि प्रतिदर्शों में

इस अनुपात में कुछ अंतर हो तो वह इतना कम होना चाहिए कि उतने या उससे अधिक अंतर के केवल संयोग से पाये जाने की प्रायिकता बहुत कम न हो। इसके विपरीत यदि इन अनुपातों में अंतर बहुत अधिक हो अर्थात् यदि टीका लगे हुए मनुष्यों में बीमारों का अनुपात उस अनुपात से बहुत कम हो जो बिना टीका लगे हुए लोगों में है—इतना कम कि यह समझना कठिन हो जाय कि यह अंतर केवल संयोगवग हो गया है—तो हम कह सकते हैं कि इन प्रेक्षणों द्वारा डाक्टरों के कथन की पुष्टि हो गयी है।

नीचे इसी प्रकार का एक उदाहरण दिया हुआ है जिससे यह स्पष्ट हो जायगा कि बड़े प्रतिदर्शों में प्रायिकता का कलन किस प्रकार किया जा सकता है।

उदाहरण (१) एक रोग भेडों में होता है जिसके कारण अधिकतर रोगी भेडों की मृत्यु हो जाती है। एक नवीन औषध का आष्कार हुआ है जिसके लिए यह दावा किया जाता है कि वह भेडों के इस रोग को ठीक कर देती है। परंतु हम यह जानते हैं कि इस विशेष रोग के अतिरिक्त भेडों की मृत्यु के अन्य भी अनेक कारण हो सकते हैं। इसके अतिरिक्त कुछ भेडें बिना किसी इलाज के भी ठीक हो सकती हैं। यह सब जानते हुए हमें इस औषध के बारे में जो दावा किया जाता है उसकी जाँच करनी है।

प्रयोग—पचास रोगी भेडों को—जो इस विशेष रोग से पीड़ित थी—यादृच्छिकीकरण द्वारा पच्चीस पच्चीस के दो कुलकों में बाँट दिया गया। हम इन कुलकों को A और B से संबोधित करेंगे। कुलक A की भेडों का इस औषध द्वारा इलाज किया गया और कुलक B की भेडों का कोई इलाज नहीं किया गया।

जब इन पचास भेडों में से प्रत्येक या तो ठीक हो गयी या मर गयी तो प्रयोग का फल निम्नलिखित था—

सारणी सख्या 94

प्रेक्षित बारम्बारता O_{ij}

	कुलक A	कुलक B	कुल
	(1)	(2)	
नीरोगी की सख्या	(1) 21	11	32
मृत्यु-सख्या	(2) 4	14	18
कुल	25	25	50

निराकरणयोग परिकल्पना H_0 औपच के वारण रोगी भेड के नीरोग होने की प्रायिकता में कुछ अतर नही पडता ।

इस परिकल्पना के आधार पर कि औपच से कुछ लाभ नहीं होता, भेड के नीरोग होने की प्रायिकता का प्राबिकलन स्पष्टतया $\frac{16}{25}$ है । इस प्रायिकता के अनुसार ऊपर की सारणी के विभिन्न खानों में प्रत्याशित सख्याएँ निम्नलिखित होगी—

सारणी संख्या 95

विभिन्न खानों में प्रत्याशित सख्याएँ E_{ij}

		कुलक A	कुलक B	कुल
		(1)	(2)	
नीरोगी की सख्या	(1)	16	16	32
	(2)	9	9	18
कुल		25	25	50

अस्वीकृति क्षेत्र—यह आपने देखा ही होगा कि इस सारणी में एक पार्श्वीय बार-बारताओं के योग 25, 25, 32 और 18 निश्चित हैं । इस कारण यदि मध्य के चार खानों में से किसी एक में सख्या दे रखी हो तो अन्य तीन खानों की सख्याओं का परिकलन किया जा सकता है । प्रत्येक खाने के लिए अलग-अलग प्रत्याशित सख्या का कलन आवश्यक नहीं है । (1, 1) खाने में 16 लिखते ही (1, 2) खाने में $32 - 16 = 16$, (2, 1) खाने में $25 - 16 = 9$ और (2, 2) खाने में $18 - 9 = 9$ लिखा जा सकता है । इस प्रकार प्रयोग के फल में केवल एक खाने में सख्या निश्चित करने की स्वतंत्रता है । अन्य खानों की सख्या का परिकलन इसी आधार पर किया जा सकता है । इस स्थिति में यह दिखाया जा सकता है कि

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

का वटन लगभग χ_1^2 है। यहाँ O_{ij} से तात्पर्य (i, j) खाने में प्रेषित सख्या से तथा E_{ij} से इसी खाने में प्रत्याक्षित सख्या से है।

इस प्रकार यदि परिकल्पित χ^2 का मान χ_1^2 के पाँच प्रतिशत बिंदु 3.841 से अधिक होगा तो हम इस परिकल्पना H_0 को अस्वीकार कर देंगे। (देखिए सारणी सख्या 9.8)

$$\begin{aligned} \text{विवलेपण—} \chi^2 &= \sum_{i=1}^2 \sum_{j=1}^3 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \\ &= \frac{5^2}{16} + \frac{5^2}{16} + \frac{5^2}{9} + \frac{5^2}{9} \\ &= 50 \left[\frac{1}{16} + \frac{1}{9} \right] \\ &= \frac{50 \times 25}{16 \times 9} \\ &= 8.68 \end{aligned}$$

निष्कर्ष—क्योंकि χ^2 का प्रेषित मान 3.841 से अधिक है इसलिए हम H_0 को अस्वीकार करते हैं। इस प्रकार हम देखते हैं कि यह प्रयोग जीपथ के बारे में किये हुए दावे की पुष्टि करता है।

उदाहरण (२) $k \times r$ वर्गीकरण

समष्टि को दो भागों में बाँटने के बजाय उसे अनेक भागों में बाँटा जा सकता है। उदाहरण के लिए (A) वे व्यक्ति जो न पढ़ सकते हैं और न लिख सकते हैं, (B) वे व्यक्ति हैं जो पढ़ तो सकते हैं, पर लिख नहीं सकते, (C) वे व्यक्ति जो पढ़ना और लिखना दोनों ही जानते हैं। यह भारत की जनता को तीन भागों में बाँटने का एक तरीका हो सकता है।

भारत की जनता को एक और प्रकार से पाँच भागों में विभाजित किया जा सकता है।

(α) वे व्यक्ति जो कांग्रेस पार्टी के अनुयायी हैं।

(β) वे व्यक्ति जो कम्युनिस्ट पार्टी के अनुयायी हैं।

- (१) वे व्यक्ति जो प्रजा सोशलिस्ट पार्टी के अनुयायी हैं ।
 (४) वे व्यक्ति जो इन तीन पार्टियों के अतिरिक्त किसी अन्य पार्टी के अनुयायी हैं ।
 (६) वे व्यक्ति जो राजनीति में बिल्कुल दिलचस्पी नहीं लेते या जिन्हें कुछ दिलचस्पी है भी तो वे किसी मौजूदा पार्टी के अनुयायी नहीं हैं ।

इन दो प्रकार के विभाजनों के संयोग से कुल 3×5 खानों में जनता के किसी भी मनुष्य को रखा जा सकता है । यदि यादृच्छिकीकरण द्वारा चुने हुए व्यक्ति के इनमें से किसी एक में होने की प्रायिकता उन दो विभाजनों में होने की प्रायिकताओं का गुणनफल हो जिनके संयोग से यह बना है, तो इस प्रकार के विभाजनों को एक दूसरे से स्वतंत्र समझा जाता है । उदाहरण के लिए यदि ऊपर के विभाजन स्वतंत्र हो तो इस घटना की प्रायिकता कि यादृच्छिकीकरण द्वारा चुना हुआ एक व्यक्ति लिखना पढ़ना नहीं जानता और उसे राजनीति में कुछ दिलचस्पी नहीं है निम्नलिखित दो घटनाओं की प्रायिकताओं का गुणनफल है । एक तो यह कि इस व्यक्ति को पढ़ना लिखना नहीं आता और दूसरी यह कि इसको राजनीति में दिलचस्पी नहीं है ।

इन गुणों की स्वतंत्रता की परिकल्पना के परीक्षण के लिए भी χ^2 -वटन का प्रयोग होता है । यदि एक प्रकार के कुल गुणों की संख्या k हो और दूसरी प्रकार के कुल गुणों की संख्या r हो तो हमें एक $k \times r$ खानों की सारणी मिलती है । ऊपर के उदाहरण में हमें एक 3×5 सारणी प्राप्त होती है जिसे नीचे दिया हुआ है । विभिन्न खानों में व्यक्ति के पाये जाने की प्रायिकता या प्रतिदर्श में विभिन्न खानों में प्रत्याशित संख्या को मालूम करने के लिए यह आवश्यक है कि हमें एक-पार्श्वीय प्रायिकताओं का ज्ञान हो । इन प्रायिकताओं का प्राक्कलन पिछले उदाहरण की भाँति एक पार्श्वीय संख्याओं के जोड़ों में कुल प्रतिदर्श परिमाण का भाग लेकर किया जाता है ।

हम प्रयोग के केवल उन फलों पर विचार कर रहे हैं जिनमें ये एक-पार्श्वीय जोड़ अचर रहते हैं जैसा इस विशेष प्रयोग में है । इस कारण किसी पंक्ति के $(k-1)$ खानों में संख्याओं का ज्ञान होने से हम बाकी एक खाने की संख्या मालूम कर सकते हैं । इसी प्रकार यदि किसी स्तंभ की $(r-1)$ संख्याएँ हमें ज्ञात हो तो बाकी एक का परिकलन किया जा सकता है । इस प्रकार यदि हमें $(k-1)$ $(r-1)$ संख्याओं का ज्ञान हो तो सारणी को पूरा किया जा सकता है । साधारण नियम द्वारा प्राप्त χ^2 -वटन परिकल्पना के अंतर्गत लगभग $\chi^2_{(k-1)(r-1)}$ वटन के बराबर होता है ।

सारणी संख्या 96

व्यक्ति के पढ़ाई के स्तर और राजनीतिक झुकाव की स्वतंत्रता की जाँच के लिए प्रेषित बारम्बारताएँ O_{ij}

i	J	α	β	γ	δ	ϵ	कुल
A		32	26	15	7	24	104
B		91	12	15	9	77	204
C		47	18	11	14	102	192
कुल		170	56	41	30	203	500

$$P(A) = \frac{104}{500} \quad P(B) = \frac{204}{500} \quad P(C) = \frac{192}{500}$$

$$P(\alpha) = \frac{170}{500} \quad P(\beta) = \frac{56}{500} \quad P(\gamma) = \frac{41}{500}$$

$$P(\delta) = \frac{30}{500} \quad P(\epsilon) = \frac{203}{500}$$

सारणी संख्या 97

गुणों की स्वतंत्रता के आधार पर ऊपर के प्रयोग में प्रत्याशित बारम्बारताएँ

$$E_{ij} = NP(i)P(j)$$

$i \backslash J$	α	β	γ	δ	ϵ	कुल
A	35.360	11.648	8.528	6.240	42.224	104.000
B	69.360	22.848	16.728	12.240	82.824	204.000
C	65.280	21.504	15.744	11.520	77.952	192.000
कुल	170.000	56.000	41.000	30.000	203.000	500.000

अस्वीकृति क्षेत्र—यदि χ^2 का परिकल्पित मान χ^2_{α} के पाँच प्रतिगत बिंदु 15 507 से अधिक होगा तो हम निराकरण योग्य परिकल्पना को अस्वीकार कर देंगे। (देखिए सारणी सख्या 9.8)

विश्लेषण —

$$\begin{aligned} \chi^2 &= \sum_{i=1}^3 \sum_{j=1}^3 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \\ &= \frac{(-3\ 360)^2}{35\ 360} + \frac{(21\ 640)^2}{69\ 360} + \frac{(-18\ 280)^2}{65\ 280} \\ &+ \frac{(14\ 352)^2}{11\ 648} + \frac{(-10\ 848)^2}{22\ 848} + \frac{(-3\ 504)^2}{21\ 540} \\ &+ \frac{(6\ 472)^2}{8\ 528} + \frac{(-1\ 728)^2}{16\ 728} + \frac{(-4\ 744)^2}{15\ 744} \\ &+ \frac{(0\ 760)^2}{6\ 240} + \frac{(-3\ 240)^2}{12\ 240} + \frac{(2\ 480)^2}{11\ 520} \\ &+ \frac{(-18\ 224)^2}{42\ 224} + \frac{(-5\ 824)^2}{82\ 824} + \frac{(24\ 048)^2}{77\ 952} \\ &> 15\ 507 \end{aligned}$$

निष्कर्ष—हम निराकरण योग्य परिकल्पना को अस्वीकार करते हैं—

इस प्रकार के परीक्षण को समागता-परीक्षण (test of homogeneity) भी कहते हैं। इसमें परिकल्पना यह होती है कि यदि समष्टि को एक गुण के अनुसार विभाजित किया जाय तो इन उप-समष्टियों का बंटन दूसरे गुण के अनुसार एक समान है। उदाहरण के लिए ऊपर दिये हुए प्रयोग में पढाई और राजनीतिक झुकाव में स्वातंत्र्य के अर्थ यह है कि यदि कुल जन-संख्या को राजनीतिक झुकाव के अनुसार विभाजित किया जाय तो इस प्रकार के प्रत्येक समूह में बिना पढे-लिखे, केवल पढना जाननेवाले और पढना तथा लिखना दोनों जाननेवालों का अनुपात बराबर होगा। इसको संकेत में निम्नलिखित ढंग से लिखा जा सकता है—

$$\begin{aligned} P(A/\alpha) &= P(A/\beta) = P(A/\gamma) = P(A/\delta) = P(A/\epsilon) \\ P(B/\alpha) &= P(B/\beta) = P(B/\gamma) = P(B/\delta) = P(B/\epsilon) \\ P(C/\alpha) &= P(C/\beta) = P(C/\gamma) = P(C/\delta) = P(C/\epsilon) \end{aligned}$$

यदि ये अनुपात बराबर हैं तो हम कह सकते हैं कि विभिन्न दृष्टिकोणवाले मनुष्यों के समूहों को मिला देने पर भी समष्टि पढाई की दृष्टि से ज्यों की त्यों बनी रहती है—अधिक असमाग (heterogenous) नहीं हो जाती ।

§ ९११ प्रसामान्य-वटन के प्रसरण सबधी परिकल्पना-परीक्षण में \bar{x}^2 -वटन का उपयोग

अभी तक \bar{x}^2 -वटन के जितने उपयोगों से हम परिचित हुए हैं उन सबमें यह आवश्यक था कि प्रतिदर्श परिमाण थोड़े-थोड़े रूप से बड़ा हो । यदि हमें यह ज्ञात हो कि समष्टि प्रसामान्य है तथा इस बात का परीक्षण करने की आवश्यकता नहीं है और हम केवल यह जानना चाहें कि इस समष्टि का प्रसरण σ^2 है अथवा नहीं तो भी हम \bar{x}^2 -वटन का प्रयोग करते हैं । साधारण रीति से माध्य का अनुमान लगाकर ऊपर दिये हुए \bar{x}^2 -परीक्षण द्वारा उसे जाँचा जा सकता है । परंतु जिस नवीन परीक्षण का हम वर्णन कर रहे हैं वह इस विशेष निराकरणयोग्य परिकल्पना के लिए अधिक शक्तिशाली है और उसके लिए प्रतिदर्श के बड़े होने की आवश्यकता नहीं है ।

मान लीजिए कि एक प्रसामान्य वटन का प्रसरण σ^2 है । यदि हम वटन का एक n परिमाण का प्रतिदर्श यादृच्छिकीकरण द्वारा लिया जाय जिसके मान x_1, x_2, \dots, x_n हों तो यह सिद्ध किया जा सकता है कि

$$n \frac{s^2}{\sigma^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2}$$

का वटन χ^2_{n-1} है । यहाँ \bar{x} से हम प्रतिदर्श माध्य $\frac{1}{n} \sum_{i=1}^n x_i$ को सूचित करते हैं ।

और s^2 उस प्रतिदर्श का प्रसरण है । इस प्रतिदर्शज (statistic) $n \frac{s^2}{\sigma^2}$ का वटन समष्टि के माध्य μ (म्यू) से सर्वथा स्वतंत्र है । इस कारण μ के अज्ञात होने पर भी समष्टि की प्रसरण सबधी परिकल्पना का परीक्षण इसकी सहायता से किया जा सकता है ।

उदाहरण—एक फौटरी में पीतल की छड़ें बनती हैं । पिछले वर्षों के अनुभव और प्रेक्षण द्वारा हम यह जानते हैं कि इन छड़ों की लंबाइयों का वटन प्रसामान्य है ।

एक ग्राहक को छड़ों की आवश्यकता है और वह एक हजार छड़े खरीदने के लिए तैयार है यदि इनकी लंबाई लगभग बराबर हो। उसका कहना है कि यदि इन हजार छड़ों की लंबाइयों का मानक विचलन ०.२ इंच से अधिक न हो तो वह इन्हें खरीदने को तैयार है। जब फैक्टरीवाले उसे बताते हैं कि एक हजार छड़ों के नापने और उनके मानक विचलन के कलन में बहुत समय तथा धनव्यय होगा जिसके कारण छड़ों की कीमत घटाने की आवश्यकता हो जायगी तो ग्राहक इस बात पर राजी हो जाता है कि दस छड़ों का एक यादृच्छिक प्रतिदर्श इन हजार छड़ों में से चुना जाय और उसके द्वारा इस निराकरणाय परिकल्पना की जाँच की जाय कि कुल समष्टि का मानक विचलन ०.२ इंच है। यदि प्रतिदर्श में मानक विचलन का अनुमान ०.२ इंच से कम आता है तब तो उसे कुछ एतराज होगा ही नहीं। परन्तु यदि प्रतिदर्श का मानक विचलन ०.२ इंच से इतना अधिक हुआ कि हमें निराकरणाय परिकल्पना को दो प्रतिदर्श स्तर पर अस्वीकार करना पड़े तो वह इन हजार छड़ों को नहीं लेगा।

*H*_० हजार छड़ों की समष्टि का मानक विचलन ०.२ इंच है।

अस्वीकृति क्षेत्र—यदि दस छड़ों के यादृच्छिक प्रतिदर्श से परिकल्पित $n \frac{s^2}{\sigma^2}$ का मान $\chi_{30-1}^2 = \chi_9^2$ के दो प्रतिशत बिंदु 19.०२ से अधिक हो तो ग्राहक छड़ों को लेने से इनकार कर देगा।

प्रेक्षण—यादृच्छिक प्रतिदर्श में छड़ों की लंबाइयाँ निम्नलिखित थी—

- (1) 60.4 इंच (2) 60.3 इंच (3) 60.8 इंच (4) 60.6 इंच (5) 60.9 इंच
(6) 60.6 इंच (7) 60.3 इंच (8) 60.1 इंच (9) 60.5 इंच (10) 60.7 इंच

$$\text{विश्लेषण—} \quad \sum_{i=1}^{10} x_i = 605.2 \text{ इंच}$$

$$\bar{x} = 60.52 \text{ इंच}$$

$$n \frac{s^2}{\sigma^2} = \frac{\sum_{i=1}^{10} (x_i - \bar{x})^2}{0.2^2}$$

$$= \frac{1}{0.04} [(-0.12)^2 + (-0.22)^2 + (0.28)^2 \\ + (0.08)^2 + (0.38)^2 + (0.08)^2]$$

$$\begin{aligned}
 &+ (-0.22)^2 + (-0.42)^2 + (-0.02)^2 \\
 &+ (0.18)^2] \\
 &= \frac{1}{0.04} [0.5560] \\
 &= 13.9
 \end{aligned}$$

निष्कर्ष—क्योंकि $n \frac{5^2}{0.2}$ का प्रेक्षित मान 19.679 से कम है इसलिए ग्राहक को छात्रों के समूह को खरीदने में कोई एतराज नहीं होना चाहिए।

इस उदाहरण के साथ हम χ^2 वटन के उपयोगों का वर्णन समाप्त करते हैं। इसका यह अर्थ कदापि नहीं है कि इस वटन के अन्य उपयोग नहीं हैं। वास्तव में बहुचर (multivariate) वटनों में विशेषकर बहुचर प्रसामान्य वटन से संबंधित अनेक निराकरणयोग्य परिकल्पनाओं के परीक्षण में इसका उपयोग होता है। परन्तु आप अभी तक बहुचर वटनों से परिचित नहीं हैं। इसलिए χ^2 के इस उपयोग का वर्णन इस स्थान पर करना उचित नहीं होगा।

सारणी सख्या 98

कुछ χ^2 वटनों के 5 और 1 प्रतिशत बिंदु

स्वातंत्र्य सख्या	5% बिंदु	1% बिंदु
1	3.841	6.635
2	5.991	7.879
3	7.815	11.341
4	9.488	13.277
5	11.070	15.086
6	12.592	16.812
8	15.507	20.090

विस्तृत सारणी के लिए देखिए—

“Statistical Tables for Biological, Agricultural and Medical Research” By Fisher and Yates

अध्याय १०

t-वटन

§ १० १ उपयोग

पिछले अध्याय के अंतिम उदाहरण में हमें यह मालूम था कि समष्टि प्रसामान्य है। इसके माध्य में हमें कुछ रुचि नहीं थी और न उसका ज्ञान था। हम इस समष्टि के प्रसरण से संबंधित निराकरणाय परिकल्पना की जांच करना चाहते थे। इसके विपरीत यह हो सकता है कि हमें यह पता हो कि समष्टि प्रसामान्य है, उसके प्रसरण का हमें ज्ञान न हो और हम उसके माध्य संबंधी किसी परिकल्पना की जांच करना चाहें। इस परीक्षण के लिए जिस वटन का उपयोग किया जाता है उसे t-वटन कहते हैं।

§ १० २ t-वटन का प्रसामान्य वटन और χ^2 -वटन से संबंध

आइए, देखा जाय कि इस वटन का प्रसामान्य वटन से और χ^2 -वटन से क्या संबंध है।

यदि X एक मादृच्छिक प्रसामान्य $N(0,1)$ चर हो Y एक χ_n^2 चर हो तथा X और Y स्वतंत्र हो तो X और Y का संयुक्त वटन $f_1(x,y)$ निम्नलिखित होगा।

$$f_1(x,y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-y/2}$$

यदि $Z = \sqrt{y/n}$ हो तो x और z का संयुक्त वटन

$$f_2(x,z) = \sqrt{\frac{n}{\pi}} \left(\frac{n}{2}\right)^{\frac{n}{2}} \frac{1}{\Gamma(\frac{n}{2})} z^{n-1} e^{-\frac{x^2 + nz^2}{2}} \quad (10.1)$$

क्योंकि हमें X और Z का संयुक्त वटन ज्ञात है इसलिए हम X और Z के किसी फलन का वटन भी मालूम कर सकते हैं। यह सिद्ध किया जा सकता है कि यदि

$$U = \frac{X}{Z} \text{ हो तो}$$

$$P[U \leq x] = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \int_{-\infty}^x \frac{du}{\left(1 + \frac{u^2}{n}\right)^{\frac{n+1}{2}}}$$

इससे संबंधित U का घनत्व-फलन स्पष्टतया निम्नलिखित है—

$$f(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \dots (10.2)$$

यह घनत्व-फलन अथवा उसका ऊपर दिया हुआ सचयी बारंबारता फलन जिस वटन को निरूपित करता है वह n स्वातन्त्र्य-राख्यावाला t -वटन कहलाता है। इसको संक्षेप में t_n -वटन कहते हैं।

§ १०.३ परिकल्पना परीक्षण

यदि एक प्रसामान्य वटन $N(\mu, \sigma)$ में से n परिमाण का एक यादृच्छिक प्रतिदर्श चुना जाय जिसमें चर के प्रेक्षित मान x_1, x_2, \dots, x_n हों तो यह हम पहिले ही बेल चुके हैं कि $\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ एक प्रसामान्य $N(0, 1)$ चर होता है, जहाँ

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

यह भी आपको पता ही है कि $\frac{s^2}{\sigma^2}$ एक χ_{n-1}^2 चर है जहाँ

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

एक दूसरे से स्वतंत्र चर हैं। इसलिए $\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim \sqrt{\frac{ns^2}{(n-1)\sigma^2}}$ एक t_{n-1} चर

है। इसमें σ/\sqrt{n} कट जाता है और हम देखते हैं कि $\frac{\bar{x} - \mu}{s} \sqrt{n-1}$ एक

t_{n-1} —चर है। क्योंकि यह माना $\frac{\bar{x}-\mu}{s} \sqrt{n-1}$ आधारभूत प्रसामान्य वटन के प्रसरण σ^2 से स्वतंत्र है, इसलिए σ^2 के अज्ञात होने पर हम t_{n-1} — वटन का उपयोग समष्टि के माध्य μ से भवधित निराकरणयोग परिकल्पना के परीक्षण के लिए कर सकते हैं। विभिन्न स्वातन्त्र्य-सख्यावाले t —वटनों की सारणियाँ सांख्यिकी ने बना रखी हैं क्योंकि इस वटन का प्रयोग परिकल्पना परीक्षण में बहुत अधिक प्रचलित है। जैसे-जैसे t —वटन की स्वातन्त्र्य-सख्या बढ़ती जाती है वह प्रसामान्य $N(0,1)$ वटन की ओर अग्रसर होता जाता है। स्वातन्त्र्य-सख्या 30 हो जाने पर ये दोनों वटन इतने अधिक समान हो जाते हैं कि इससे अधिक किसी भी स्वातन्त्र्य-सख्या के होने पर t —वटन के स्थान पर $N(0,1)$ वटन के प्रयोग से कोई विशेष त्रुटि की सम्भावना नहीं रहती।

सारणी सख्या 10-1

कुछ t —वटनों के 5.0, 2.5, 1.0 तथा 0.5 प्रतिशत बिंदु

स्वातन्त्र्य-सख्या	12	15	18	21	24
5.0% बिंदु	1.782	1.753	1.734	1.721	1.711
2.5% बिंदु	2.179	2.131	2.101	2.080	2.064
1.0% बिंदु	2.681	2.602	2.552	2.518	2.492
0.5% बिंदु	3.055	2.947	2.878	2.831	2.797

विस्तृत सारणी के लिए देखिए—

“Statistical Tables for Biological, Agricultural and Medical Research” by Fisher and Yates

§ १०.४ उदाहरण

(१) यह कहा जाता है कि अमेरिका-निवासियों की औसत ऊंचाई छ फुट है। इस परिकल्पना की जाँच के लिए पच्चीस अमेरिका-निवासियों का एक यादृच्छिक प्रतिदर्श लिया गया और उनकी ऊँचाइयों को नापा गया। इस प्रयोग का फल निम्न-लिखित था—

$$\bar{x} = 5 \text{ फुट } 10.0 \text{ इंच}$$

$$s = 0.5 \text{ फुट } 0.5 \text{ इंच}$$



निराकरणाय परिकल्पना H_0 :

अमेरिका-वासियों की औसत ऊंचाई छ फुट है।
अस्वीकृति क्षेत्र

यदि प्रतिदर्श में ऊँचाइयों का माध्य 6 फुट से इतना कम हो कि निराकरणाय परिकल्पना के आधार पर प्रेषित अथवा उससे भी कम माध्य होने की प्रायिकता 0.5 प्रतिशत से भी कम हो अथवा यदि यह माध्य 6 फुट से इतना अधिक हो कि निराकरणाय परिकल्पना के आधार पर प्रेषित अथवा उससे भी अधिक माध्य की प्रायिकता 0.5 प्रतिशत या उससे भी कम हो तो निराकरणाय परिकल्पना को अस्वीकार कर दिया जायगा। इस प्रकार निराकरणाय परिकल्पना के सत्य होने पर भी उसको अस्वीकार करने की कुल प्रायिकता एक प्रतिशत है।

इस तरह यदि $\left| \frac{\bar{x} - 6 \text{ फुट}}{s/\sqrt{n-1}} \right|$ का मान t_{α} के 0.5 प्रतिशत बिंदु 2.797 से अधिक हो तो हम H_0 को अस्वीकार करेंगे। (देखिए सारणी सख्या 101)

विक्षेपण-
$$\left| \frac{\bar{x} - 6 \text{ फुट}}{s/\sqrt{n-1}} \right| = \frac{20}{0.5} \sqrt{24}$$

$$= 8\sqrt{6}$$

निष्कर्ष-

$\left| \frac{\bar{x} - 6 \text{ फुट}}{s/\sqrt{n-1}} \right|$ का प्रेषित मान 2.797 से बहुत अधिक है, इसलिए हमें H_0 को अस्वीकार करना होगा।

इस परिकल्पना की जाँच में हम इस अभिधारणा को लेकर चले हैं कि अमेरिका वासियों की ऊँचाइयों का वटन प्रसामान्य है। यदि यह अभिधारणा गलत हो तो ऊपरलिखित परीक्षण का सैद्धांतिक आधार ही जाता रहेगा। हम यह देख चुके हैं कि समष्टि के प्रसामान्य न होने पर भी यदि प्रतिदर्श काफी बड़ा हो तो \bar{x} का वटन लगभग प्रसामान्य होता है। इसी प्रकार देखा गया है कि यदि प्रतिदर्श बड़ा न हो तो

$\frac{\bar{x} - \mu}{s/\sqrt{n-1}}$ का वटन लगभग t_{n-1} होता है। इस कारण समष्टि के प्रसामान्य न होने पर भी t_{n-1} वटन के प्रयोग से जाँच में विशेष त्रुटि नहीं होती।

§ १०.५ एक तरफा और दो तरफा परीक्षण

ऊपर के उदाहरण में $\frac{\bar{x}-\mu}{s/\sqrt{n-1}}$ का मान 2.797 से बड़ा हो या—2.797 से

छोटा हो, इन दोनों ही अवस्थाओं में हमने H_0 को अस्वीकार करने का निश्चय किया था। इस प्रकार के परीक्षण को दो-तरफा परीक्षण (two-sided test) कहते हैं। इसके विपरीत कुछ अवस्थाएँ ऐसी हो सकती हैं जिनमें हम निराकरणाय परिकल्पना

की केवल उसी समय अस्वीकार करते हैं जब $\frac{\bar{x}-\mu}{s/\sqrt{n-1}}$ का मान बहुत बड़ा हो। बहुत

छोटा होने पर अस्वीकार नहीं करते। इसी प्रकार कुछ अन्य अवस्थाएँ ऐसी भी हो सकती हैं जिनमें निराकरणाय परिकल्पना केवल उसी समय अस्वीकार की जाती है

जब $\frac{\bar{x}-\mu}{s/\sqrt{n-1}}$ का मान बहुत छोटा हो—बहुत बड़ा होने पर नहीं। इस प्रकार के

परीक्षण को एक-तरफा परीक्षण (one-sided test) कहते हैं। आइए, अब हम एक उदाहरण द्वारा एक-तरफा परीक्षण से परिचय प्राप्त करें।

(२) एक शरीर-रचना विशेषज्ञ (anatomist) ने गहल अध्ययन के पश्चात् यह सिद्धान्त निकाला कि साधारणतया मनुष्य का दाहिना हाथ बायें हाथ से अधिक लंबा होता है।

निराकरणाय परिकल्पना H_0

दाहिने और बायें हाथों की औसत लंबाइयाँ बराबर हैं। यदि दाहिने हाथ की लंबाइयों की समष्टि का माध्य μ_1 हो और बायें हाथ की लंबाइयों की समष्टि का माध्य μ_2 हो तो

$$\mu_1 = \mu_2$$

अथवा

$$\mu_1 - \mu_2 = 0 \quad \dots\dots(10.3)$$

इसलिए निराकरणाय परिकल्पना को दूसरे शब्दों में भी रखा जा सकता है—“दाहिने और बायें हाथों की लंबाइयों के अंतर की समष्टि का माध्य शून्य है।”

वैकल्पिक परिकल्पना H_1 :

दाहिने और बायें हाथों की लंबाइयों के अंतर की समष्टि का माध्य शून्य से अधिक है।

$$\mu_1 - \mu_2 > 0$$

यही वह सिद्धांत है जो शरीर रचना विशेषज्ञ ने निकाला है।

प्रयोग—परिकल्पना की जाँच के लिए 16 मनुष्यों का एक यादृच्छिक प्रतिदर्श लिया गया। इस प्रतिदर्श में चुने हुए व्यक्तियों के दाहिने और बायें हाथों की लंबाइयाँ मापी गयीं।

यदि दाहिने हाथ की लंबाइयों के प्रतिदर्श-माध्य को \bar{x}_1 तथा बायें हाथ की लंबाइयों के प्रतिदर्श माध्य को \bar{x}_2 से सूचित किया जाय, प्रतिदर्श के 1-वें मनुष्य के दाहिने और बायें हाथ की लंबाइयों को क्रमशः x_{11} तथा x_{21} से सूचित किया जाय तो इस प्रयोग के फल को निम्नलिखित रूप में रखा जा सकता है।

$$\bar{x}_1 = 2 \text{ फुट } 10 \text{ इंच}$$

$$\bar{x}_2 = 2 \text{ फुट } 05 \text{ इंच}$$

$$\begin{aligned} s^2 &= \frac{1}{16} \sum_{i=1}^{16} \left\{ (x_{1i} - x_{2i}) - (\bar{x}_1 - \bar{x}_2) \right\}^2 \\ &= \frac{1}{16} \left\{ \sum_{i=1}^{16} (x_{1i} - x_{2i})^2 - 16(\bar{x}_1 - \bar{x}_2)^2 \right\} \\ &= 0.52 \text{ वर्ग इंच} \end{aligned}$$

$$\therefore s = 0.7141 \text{ इंच}$$

अन्वोक्रुति क्षेत्र

$$\text{यदि } \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{15}} = \frac{(\bar{x}_1 - \bar{x}_2)\sqrt{15}}{s} \text{ का मान } t_{15} \text{ के पाँच प्रतिशत बिंदु}$$

1.753 से अधिक होगा तो निराकरणिय परिकल्पना H_0 को अस्वीकार करके हम परिकल्पना H_1 को स्वीकार करेंगे। (देखिए सारणी सख्या 10.1)

$$\text{विश्लेषण } \left(\frac{\bar{x}_1 - \bar{x}_2}{s} \right) \frac{\sqrt{15}}{5} = \frac{0.50 \times 3.87}{0.7141} > 1.753$$

निष्कर्ष

दोनों हाथों की लंबाईयाँ बराबर होने की परिकल्पना को अस्वीकार करके हम कह सकते हैं कि प्रयोग का फल शरीर-रचना विशेषज्ञ के सिद्धान्त के अनुकूल है।

इस उदाहरण में हमने एक-तरफा परीक्षण का उपयोग किया है। इसमें निराकरणिय परिकल्पना के सत्य होने पर भी उसको अस्वीकार करने की प्रायिकता पाँच प्रतिशत है। हम इसमें प्रेक्षित मान की तुलना t -वटन के पाँच प्रतिशत बिंदु से करते हैं। यदि हम दो-तरफा परीक्षण का प्रयोग करते तो प्रेक्षित मान की तुलना

t -वटन के 2.5 प्रतिशत बिंदु में की जाती। यदि $\frac{\bar{x} - \mu}{s/\sqrt{n-1}}$ का घनात्मक मान

इस बिंदु से अधिक होता तो निराकरणिय परिकल्पना को अस्वीकार कर दिया जाता। निराकरणिय परिकल्पना के सत्य होते हुए भी उसे अस्वीकार करने की प्रायिकता तब भी पाँच प्रतिशत ही होती। t -वटन की भाँति प्रसामान्य वटन के उपयोग में भी परिस्थिति के अनुसार एक-तरफा अथवा दो-तरफा परीक्षण होता है।

१०.६ द्वि-प्रतिदर्श परीक्षण (two sample test)

पिछले उदाहरण में आपने दो समष्टियों के माध्यों के बराबर होने की परिकल्पना की जाँच की थी, परंतु इसकी आवश्यकता नहीं थी कि दोनों समष्टियों में से प्रतिदर्शों का अलग-अलग चुनाव करें, क्योंकि एक ही मनुष्य से दोनों समष्टियों का माप लिया जा सकता था। परंतु ऐसी कई स्थितियाँ हो सकती हैं जिनमें दोनों समष्टियों में से अलग-अलग प्रतिदर्श चुनने की आवश्यकता हो।

यदि एक समष्टि में से n_1 परिमाण का और दूसरी में से n_2 परिमाण का प्रतिदर्श यादृच्छिकीकरण द्वारा स्वतंत्र रूप से चुना जाय, इन प्रतिदर्शों के माध्य क्रमशः

\bar{x}_1 तथा \bar{x}_2 हो और दोनों समष्टियों में प्रसरण बराबर हो तो

$$\begin{aligned} V(\bar{x}_1 - \bar{x}_2) &= E[(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)]^2 \\ &= E[(\bar{x}_1 - \mu_1) - (\bar{x}_2 - \mu_2)]^2 \\ &= E[(\bar{x}_1 - \mu_1)^2 + (\bar{x}_2 - \mu_2)^2 - 2(\bar{x}_1 - \mu_1)(\bar{x}_2 - \mu_2)] \\ &= E(\bar{x}_1 - \mu_1)^2 + E(\bar{x}_2 - \mu_2)^2 - 2E(\bar{x}_1 - \mu_1)E(\bar{x}_2 - \mu_2) \\ &= \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} - 2 \times 0 \times 0 \\ &= \sigma^2 \left[\frac{1}{n_1} + \frac{1}{n_2} \right] \end{aligned}$$

जहाँ σ^2 दोनों समष्टियों का प्रसरण है। प्रतिदर्श माध्यों के अंतर के इस प्रसरण का निम्नलिखित प्राक्कलन है

$$\hat{V}(\bar{x}_1 - \bar{x}_2) = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} \times \left[\frac{1}{n_1} + \frac{1}{n_2} \right]$$

$$\text{जहाँ } n_1 s_1^2 = \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2$$

$$n_2 s_2^2 = \sum_{i=1}^{n_2} (x_{2i} - \bar{x}_2)^2$$

यहाँ पहिले प्रतिदर्श की i -वीं इकाई के मान को x_{1i} तथा दूसरे प्रतिदर्श के i -वीं इकाई के मान को x_{2i} से सूचित किया गया है।

एक प्रतिदर्श परीक्षण में $\bar{x} - \mu$ को उसके मानक विचलन के अनुमान $\sqrt{\frac{s^2}{n-1}}$

से विभाजित करने पर जो राशि प्राप्त होती थी वह एक t_{n-1} चर थी। उसी प्रकार द्वि-प्रतिदर्श परीक्षण में $(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)$ को उसके मानक विचलन के प्राक्कलन द्वारा विभाजित करने से हमें जो चर प्राप्त होता है उसका वॉटन $t_{n_1+n_2-2}$ है। यदि परिकल्पना यह हो कि दोनों समष्टियों के माध्य बराबर हैं तो $\mu_1 - \mu_2 = 0$ । इसलिए इस परिकल्पना के अंतर्गत

$$t_{n_1+n_2-2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left[\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} \right] \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}}$$

$$= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

आइए, अब एक उदाहरण की सहायता से हम इस परीक्षण से भली-भाँति परिचित हो जायें।

§ १० ७ उदाहरण

गन्ने की दो किस्में हैं—एक भारतीय और दूसरी जावा की। यह कहा जाता है कि भारतीय गन्ने की अपेक्षा जावा के गन्ने में चीनी की मात्रा अधिक है। इस परिकल्पना की जाँच के लिए दोनों प्रकार के गन्नों के दस दस गट्टर चुने गये और उनको दबाकर रस निकाल कर उनमें चीनी का अनुपात मालूम किया गया।

निराकरणिय परिकल्पना H_0

इन दोनों प्रकार के गन्नों में औसतन चीनी का अनुपात बराबर है।

वैकल्पिक परिकल्पना H_1

औसतन जावा के गन्नों में चीनी की मात्रा अधिक है।

अस्वीकृति क्षेत्र

$$\text{यदि } t = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{10s_1^2 + 10s_2^2}} \sqrt{\frac{10 \times 10 \times (10 + 10 - 2)}{10 + 10}}$$

$$= \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{s_1^2 + s_2^2}} \times 3$$

का प्रेक्षित मान t_{18} के पाँच प्रतिशत बिंदु 1.734 से अधिक होगा तो वैकल्पिक परिकल्पना की तुलना में निराकरणिय परिकल्पना को अस्वीकार किया जायगा (देखिए सारणी सख्या 10.1)

प्रेक्षण—गन्ने के विभिन्न गट्टरों से प्राप्त चीनी की मात्रा (पीण्ड में) नीचे की सारणी में दी गयी है।

सारणी संख्या 10.2

भारतीय गन्ना		जावा का गन्ना	
गट्टर संख्या	चीनी की मात्रा	गट्टर संख्या	चीनी की मात्रा
(1)	(2)	(3)	(4)
1	15	1	21
2	19	2	18
3	21	3	16
4	17	5	20
5	19	5	23
6	16	6	16
7	15	7	19
8	22	8	20
9	17	9	23
10	20	10	17
कुल	181	कुल	293

वितरण

$$\bar{x}_1 = 18.1$$

$$\bar{x}_2 = 19.3$$

$$\sum_{i=1}^{10} x_{1i}^2 = 3331$$

$$\sum_{i=1}^{10} x_{2i}^2 = 3785$$

$$\begin{aligned}\therefore 10s_1^2 &= \sum_{i=1}^{10} x_1^2 - 10\bar{x}_1^2 \\ &= 3331 - 3276 \\ &= 549\end{aligned}$$

$$\begin{aligned}10s_2^2 &= \sum_{i=1}^{10} x_2^2 - 10\bar{x}_2^2 \\ &= 3785 - 3724 \\ &= 609\end{aligned}$$

$$\begin{aligned}t &= \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{s_1^2 + s_2^2}} \times \sqrt{3} \\ &= \frac{12 \times 3}{\sqrt{115/10}} \\ &= \frac{36}{3.39} \\ &< 1.734\end{aligned}$$

निष्कर्ष—क्योंकि निकष (criterion) का प्रेक्षित मान 1.734 से कम है, इसलिए इस प्रयोग के आधार पर निराकरणयोग्य परिकल्पना को अस्वीकार करने का कोई कारण नहीं है।

इस उदाहरण में हमने एक तरफा परीक्षण का प्रयोग किया है। परंतु जिस प्रकार एक प्रतिदर्श के लिए दो तरफा परीक्षण होता है उसी प्रकार वैकल्पिक परिकल्पना के किसी विशेष दिशा में झुकाव न होने पर द्वि-प्रतिदर्श के लिए भी दो-तरफा परीक्षण का उपयोग किया जाता है।

§ १०८—परीक्षण पर प्रतिबन्ध

यह ध्यान देने योग्य बात है कि इस परीक्षण का आधार यह अभिधारणा है कि दोनों समष्टियों के प्रसरण समान हैं। यदि प्रसरण बहुत भिन्न हों तो इस परीक्षण का

उपयोग युक्तियुक्त नहीं है। यह स्वाभाविक है कि आप जानना चाहें कि दोनों समष्टियों के प्रसरण बराबर है या नहीं। यह किस प्रकार मालूम किया जाय ? 'दो प्रभामान्य वटनों के प्रसरण बराबर हैं' इस निराकरणीय परिकल्पना की परीक्षा करने के साधन वास्तव में सांख्यिकों के पास हैं। बिना इस प्रकार के परीक्षण के अथवा बिना लंबे अनुभव के इस अभिवारणा को कोई भी वैज्ञानिक मानने की तैयार नहीं होगा। आपका यह सोचना ठीक है कि इस अभिवारणा का परीक्षण पहले और I-परीक्षण का प्रयोग बाद में होना चाहिए।

इस नये परीक्षण के लिए हमें एक नवोन प्रकार के वटन का उपयोग करना पड़ता है जिसे M-वटन कहते हैं। इसका और इसके उपयोग का सक्षिप्त वर्गन अगले अध्याय में दिया गया है।

अध्याय ११

F-वंटन

§ ११.१ F-वंटन और χ^2 -वंटन का सम्बन्ध

मान लीजिए कि X और Y दो यादृच्छिक चर हैं। X का वंटन $\chi_{n_1}^2$ तथा Y का वंटन $\chi_{n_2}^2$ है। तब $F = \frac{X}{n_1} - \frac{Y}{n_2}$ का घनत्व-फलन $f(x)$ निम्नलिखित है—

$$f(x) = \left[\frac{n_1}{n_2} \right]^{\frac{n_1}{2}} \frac{\Gamma \left[\frac{n_1 + n_2}{2} \right]}{\Gamma \left[\frac{n_1}{2} \right] \Gamma \left[\frac{n_2}{2} \right]} \frac{x^{\frac{n_1}{2} - 1}}{\left[1 + \frac{n_1 x}{n_2} \right]^{\frac{n_1 + n_2}{2}}} \dots \quad (11.1)$$

इस वंटन को n_1 तथा n_2 स्वातन्त्र्य-संख्याओं का F-वंटन कहते हैं। संक्षेप में इसे F_{n_1, n_2} से भी सूचित करते हैं। इस वंटन का प्रयोग बहुत अधिक होने के कारण, सांख्यिकी ने विभिन्न स्वातन्त्र्य-संख्याओं के F-वंटनों के प्रतिशतता-बिंदुओं की सारणी तैयार कर रखी है।

सारणी संख्या 11 1

कुछ F-वंटनों के 5 और 1 प्रतिशत बिंदु

वंटन	5% बिंदु	1% बिंदु
$F_{3,6}$	4.76	9.78
$F_{3,15}$	3.29	5.42
$F_{3,21}$	3.07	4.87
$F_{4,11}$	3.36	5.67
$F_{6,15}$	2.90	4.56
$F_{7,21}$	2.48	3.64
$F_{9,6}$	3.19	5.36

विस्तृत सारणी के लिए देखिए

“Statistical Tables for Biological, Agricultural and Medical Research” by Fisher and Yates

§ ११२ परिकल्पना परीक्षण

मान लीजिए कि दो प्रसामान्य समष्टियाँ हैं जिनके माध्य क्रमशः μ_1 और μ_2 तथा प्रसरण क्रमशः σ_1^2 और σ_2^2 हैं। इन दो समष्टियों में से क्रमशः n_1 तथा n_2 परिमाण के यादृच्छिक प्रतिदर्श स्वतंत्र रूप से चुने जाते हैं। इन प्रतिदर्शों के प्रसरण क्रमशः s_1^2 और s_2^2 हैं।

अतः $\frac{n_1 s_1^2}{\sigma_1^2}$ एक $\chi_{n_1-1}^2$ चर है

तथा $\frac{n_2 s_2^2}{\sigma_2^2}$ एक $\chi_{n_2-1}^2$ चर है।

ये दोनों चर एक दूसरे से स्वतंत्र भी हैं। इसलिए

$$F = \frac{n_1 s_1^2}{(n_1 - 1)\sigma_1^2} \div \frac{n_2 s_2^2}{(n_2 - 1)\sigma_2^2} \text{ एक } F_{n_1-1, n_2-1} \text{ चर है।}$$

यदि निराकरणीय परिकल्पना यह है कि $\sigma_1^2 = \sigma_2^2$ तो इसके अन्तर्गत

$$F = \frac{n_1 s_1^2 / n_1 - 1}{n_2 s_2^2 / n_2 - 1} \text{ एक } F_{n_1-1, n_2-1} \text{ चर है। इस गुण का प्रयोग परि-}$$

कल्पना की परीक्षा के लिए सरलता से किया जा सकता है। यदि प्रयोग में प्रेक्षित F का मान F_{n_1-1, n_2-1} के एक पूर्व निश्चित प्रतिशतता बिन्दु से अधिक हो तो हम निराकरणीय परिकल्पना को अस्वीकार करते हैं। यदि इस परिकल्पना को अस्वीकार किया जाता है तो द्वि-प्रतिदर्शीय t -परीक्षण युक्ति-संगत नहीं है। यदि परीक्षण द्वारा परिकल्पना को अस्वीकार नहीं किया जाता तो इसका यह अर्थ नहीं है कि उसकी सत्यता सिद्ध हो गयी। इसका अर्थ केवल इतना ही है कि प्रयोग के फल परिकल्पना के सत्य होने की स्थिति में काफी संभव थे और इस कारण वे परिकल्पना के विरुद्ध कोई साक्ष्य नहीं देते।

§ ११३ उदाहरण

आइए, अब यह देखा जाय कि इसका उपयोग पिछले उदाहरण में किस प्रकार किया जा सकता है।

निराकरणीय परिकल्पना H_0

भारतीय और जावा द्वीपीय गध्रों में चीनी के वंटनों के प्रसरण बराबर हैं।

वैकल्पिक परिकल्पना H_1

ये प्रसरण बराबर नहीं हैं।

अस्वीकृति क्षेत्र

यदि $F = \frac{10s_2^2/9}{10s_1^2/9} = \frac{s_2^2}{s_1^2}$ का प्रेक्षित मान $F_{9,9}$ के पाँच प्रतिशत बिंदु

3.19 से अधिक हो तो वैकल्पिक परिकल्पना की तुलना में निराकरणीय परिकल्पना को अस्वीकार कर दिया जायगा। (देखिए सारणी सख्या IX.1)

विश्लेषण

प्रयोग के प्रेक्षणों के अनुसार

$$F = \frac{60.1}{54.9} < 3.19$$

निष्कर्ष—प्रेक्षणों के आधार पर हम निराकरणीय परिकल्पना को अस्वीकार नहीं कर सकते।

प्रयोग-विश्लेषण में F -वटन का उपयोग बहुत अधिक होता है। इसका वर्णन उन अन्य अध्यायों में दिया हुआ है जिनका संबंध प्रयोग-अभिकल्पना और प्रयोग-विश्लेषण से है। इस ऊपर के उदाहरण के साथ हम परिकल्पना की जाँच के उदाहरणों और साधारण परिचय को समाप्त करते हैं और अब हम अगले अध्याय में परिकल्पना की जाँच के साधारण सिद्धांतों का अध्ययन करेंगे।

परिकल्पना की जाँच के साधारण सिद्धान्त

§ १२.१ जाँच की परिचित विधि की आलोचना

अब तक परिकल्पना की जाँच की मनोवैज्ञानिक पृष्ठभूमि को आप भली-भाँति समझ गये होंगे। हम पहिले किसी प्रतिदर्शज (statistic) की स्थापना करते हैं जिसके मान के आधार पर हम परिकल्पना को स्वीकार अथवा अस्वीकार करेंगे। इस प्रतिदर्शज को परिकल्पना-परीक्षण का निकष (criterion) कहा जाता है। आपमें से कुछ लोगो को यह विचित्र लगा होगा कि इस जाँच के लिए हम इस निकष के प्रेक्षित मान की प्रायिकता का कलन नहीं करते, किन्तु इस घटना की प्रायिकता का कलन करते हैं कि निकष का मान या तो उपर्युक्त प्रेक्षित मान के बराबर हो अथवा उससे भी अधिक हो। कदाचित् आप अस्पष्ट रूप से इस तरीके के आधार को समझते हों। परन्तु कुछ पाठक ऐसे भी हो सकते हैं जिन्हें साक्ष्यिक पर सदेह हो कि वह मानबुझकार केवल आसानी के लिए ही इस प्रकार से प्रायिकता का कलन करता है तथा इसमें युक्ति कुछ भी नहीं है।

फिर भी यह तो स्पष्ट ही है कि किसी भी सतत वटन में, उदाहरण के लिए एक प्रसामान्य वटन में, किसी विशेष मान के प्रेक्षण की प्रायिकता शून्य है। असतत वटन में भी यदि चर सँकड़ो मान धारण कर सकता हो तो किसी भी विशेष मान को धारण करने की प्रायिकता बहुत छोटी हो सकती है। इस कारण केवल प्रेक्षित घटना की प्रायिकता के छोटे होने पर यदि हम निराकरणिय परिकल्पना को अस्वीकार करने का निर्णय करें तो प्रयोग करने की कोई आवश्यकता ही नहीं है। क्योंकि यह स्पष्ट है कि चाहे प्रयोग का फल कुछ भी हो उसकी प्रायिकता बहुत ही कम अथवा शून्य होगी और इस कारण हम उसको अस्वीकार कर देंगे।

§ १२.२ अस्वीकृति क्षेत्र

वास्तव में यदि हम परिकल्पना को पाँच प्रतिशत स्तर पर अस्वीकार करने का निश्चय करते हैं तो हमें एक अन्तराल अथवा भागो के एक कुलक की परिभाषा

देनी होगी जिसमें प्रेक्षित मान के पाये जाने की प्रायिकता परिकल्पना के अन्तर्गत पाँच प्रतिशत हो। इसको अस्वीकृति-क्षेत्र अथवा संशय-अंतराल (critical region) कहते हैं। यदि प्रेक्षित मान अस्वीकृति-क्षेत्र में पाया जाता है तब हम निराकरणीय परिकल्पना को अस्वीकार करते हैं, अन्यथा नहीं। इस प्रकार यदि परिकल्पना वास्तव में सत्य हो तो गलती से उसको अस्वीकार करने की प्रायिकता पाँच प्रतिशत रह जाती है।

मान लीजिए, हम प्रतिदर्श माध्य और प्रत्याशित माध्य के अन्तर को $(\bar{x} - \mu)$ से सूचित करते हैं। यदि हम अस्वीकृति-क्षेत्र को इस प्रकार चुनें कि जब $\bar{x} - \mu = 1$ हो तब तो हम परिकल्पना को अस्वीकार कर देंगे, परन्तु जब यह अन्तर बहुत अधिक हो, जैसे 3 या 4, तब हम परिकल्पना को अस्वीकार नहीं करेंगे तो यह मनोवैज्ञानिक दृष्टिकोण से अनुचित होगा। यह स्वाभाविक है कि अस्वीकृति-क्षेत्र में प्रेक्षित और प्रत्याशित मानों के अन्तरों को व्यक्त करनेवाली सख्याएँ बड़ी-बड़ी हो और यदि कोई विशेष सख्या इस क्षेत्र में विद्यमान हो तो उससे बड़ी सब सख्याएँ भी अस्वीकृति-क्षेत्र में ही हों।

§ १२'३ एक तरफा परीक्षण

यदि किसी के पास एक ऐसी वैकल्पिक परिकल्पना है जिसके अनुसार हम घनात्मक अन्तर की आशा कर सकते हैं। तब प्रश्न केवल निराकरणीय परिकल्पना की जाँच ही नहीं है। बल्कि निराकरणीय और वैकल्पिक परिकल्पनाओं में से एक का चुनाव करना है। इस प्रकार की स्थिति में स्वाभाविक है कि हम एकतरफा परीक्षण का प्रयोग करें।

§ १२'४ विभिन्न निकषों से अलग-अलग निष्कर्ष निकालने की संभावना

ऊपर लिखे तर्क कई लोगों को सतोपश्रद और यथेष्ट मालूम हो सकते हैं। फिर भी परिकल्पना परीक्षण के सिद्धान्तों का व्यवस्थित विकास आवश्यक है। एक ही प्रतिदर्श के प्रेक्षणों से ऐसे अनेक प्रतिदर्शज (statistic) बन सकते हैं जिनके वटनों को हम निराकरणीय परिकल्पना के अन्तर्गत जानते हैं। यह संभव है कि यद्यपि किसी एक प्रतिदर्शज के दृष्टिकोण से परिकल्पना को अस्वीकार किया जा सकता है परन्तु किसी दूसरे प्रतिदर्शज के विचार से उस परिकल्पना को त्यागने का कोई कारण दृष्टिगोचर न हो। ऐसी अवस्था में हमें यह जानना आवश्यक है कि किस प्रतिदर्शज के आधार पर परीक्षण करें।

एक उदाहरण के द्वारा हम ऊपर के बचन को स्पष्ट कर देना चाहते हैं। मान लीजिए कि ह्य जानते हैं कि समष्टि प्रसामान्य है और उसका मानक विचलन σ है। हम इस निराकरणयोग्य परिकल्पना का परीक्षण करना चाहते हैं कि उसका माध्य μ है। इस परिकल्पना के लिए हम एक परीक्षण का वर्णन पहले ही कर चुके हैं जिसमें प्रतिदर्श-माध्य \bar{x} और μ का अन्तर एक विशेष मान से अधिक होने पर हम परिकल्पना का त्याग करते हैं। इस परिकल्पना की जाँच का दूसरा तरीका निर्भरलिखित भी हो सकता है।

हम यह जानते हैं कि एक प्रसामान्य समष्टि के माध्य और माध्यिका बराबर होते हैं। इसलिए किसी प्रोक्षित राशि के μ से कम होने की उतनी ही प्रायिकता है जितनी μ से अधिक होने की। इसलिए परिकल्पना के अनुसार यह आशा की जाती है कि प्रतिदर्श में जितनी राशियाँ μ से छोटी होंगी उतनी ही μ से बड़ी भी होंगी। इस कारण μ से बड़ी राशियों की संख्या बहुत अधिक होने पर अथवा बहुत कम होने पर भी हम परिकल्पना का त्याग कर सकते हैं। इस प्रकार प्रसामान्य बटन के माध्य के μ होने के लिए ऊपर लिखे दो परीक्षण हो सकते हैं जो एक-दूसरे से भिन्न हैं। हो सकता है कि एक के अनुसार परिकल्पना अस्वीकृत हो और दूसरी के अनुसार नहीं हो। उदाहरण के लिए यदि

$$\begin{array}{ll} \sigma=2 & \mu=5 \\ \bar{x}=4 & n=25 \\ n_1=15 & n_2=10 \end{array}$$

जहाँ \bar{x} प्रतिदर्श माध्य और n प्रतिदर्श परिमाण है। n_1 उन प्रेक्षणों की संख्या है जिनके मान $\mu=5$ से कम हैं तथा n_2 उन प्रेक्षणों की संख्या है जिनके मान 5 से अधिक हैं। जिस द्विपद बटन के प्राचल 25 और $\frac{1}{2}$ हो उसके द्वारा n_2 के 15 या इससे भी अधिक होने की प्रायिकता का कलन किया जा सकता है।

अपूर्ण B -फलन सारणी के अनुसार यह प्रायिकता 0.2121781 है। यह इतनी अधिक है कि इसके आधार पर परिकल्पना को अस्वीकार करना समभव नहीं है।

किन्तु दूसरी ओर हमें पता है कि परिकल्पना के अन्तर्गत $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ का बटन $N(0, 1)$

है, इसलिए परिकल्पना-परीक्षण $t = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ को निकट मानकर भी किया जा सकता

है। किन्तु दूसरी ओर हमें पता है कि परिकल्पना के अन्तर्गत $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ का बंटन

$N(0,1)$ है इसलिये परिकल्पना-परीक्षण $t = \left| \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \right|$ को निकप मानकर भी

किया जा सकता है। और प्रयोग में $t = \frac{1}{2.5}$

$$= 2.5$$

प्रसामान्य बंटन के अनुसार निकप t के 2.5 अथवा उससे भी अधिक होने की प्रायिकता 5% से कम है। इस कारण हम प्रसामान्य समष्टि के माध्य के मान के 5 होने को अस्वीकार करते हैं।

इस प्रकार एक ही प्रतिदर्श पर निर्भर दो परीक्षणों के नतीजे अलग-अलग होना संभव है। इस दृश्य में यह जानना आवश्यक है कि निर्णय किस परीक्षण पर आधारित होना चाहिए। यह स्पष्ट है कि यदि हम 5% के स्तर पर परीक्षण करते हैं तो परिकल्पना के सत्य होते हुए भी उसके अस्वीकार किये जाने की त्रुटि की प्रायिकता हर एक परीक्षण के लिए समान होगी। इसलिए इस प्रकार की त्रुटि के कम या अधिक होने को हम परीक्षण चुनने के लिए निकप (criterion) नहीं मान सकते। नीमन और पीयरसन (Neyman and Pearson) ने इसके लिए एक अन्य निकप का प्रतिपादन किया है तथा उसके ऊपर परिकल्पना-परीक्षण के सिद्धान्तों का एक ढाँचा खड़ा किया है। इसका वर्णन आगे के कुछ पृष्ठों में किया गया है। परन्तु प्रो० रोनाल्ड फिशर और उनके अनुयायियों की एक अन्य विचारधारा है जिसके अनुसार वैज्ञानिक अध्ययन में नीमन और पीयरसन द्वारा प्रतिपादित विचार-मद्धति युक्तिसंगत नहीं है। इसलिए प्रो० फिशर की विचारधारा का भी संक्षेप में वर्णन किया जायेगा।

§ १२५ नीमन-पीयरसन सिद्धान्त

नीमन-पीयरसन सिद्धान्त का आरम्भ इस प्रेरण से होता है कि किसी भी परिकल्पना-परीक्षण के उपयोग में दो प्रकार की त्रुटियाँ संभव हैं। उनके अनुसार परीक्षण के अंत में दो ही फल हो सकते हैं। या तो हम परिकल्पना को स्वीकार करें अथवा अस्वीकार कर दें। यदि परिकल्पना सत्य न हो और हम उसे स्वीकार कर लें अथवा वह सत्य हो और हम उसे अस्वीकार कर दें—इन दोनों ही स्थितियों में हम भूल करते

है। इनको सिद्धान्त में क्रमबद्ध दूसरी और पहली किस्म की त्रुटि (errors of second and first kind) कहते हैं।

§ १२'५'१ पहली प्रकार की त्रुटि—परिकल्पना को अस्वीकार करने की भूल जब वह वास्तव में सत्य है।

§ १२'५'२ दूसरी प्रकार की त्रुटि—परिकल्पना को स्वीकार करने की भूल जब कि वह वास्तव में असत्य है।

यदि कोई परीक्षण दोनों प्रकार की त्रुटियों की प्रायिकता को अधिक से अधिक घटा सके तो उसको दूसरे परीक्षणों की अपेक्षा अच्छा समझा जायेगा। यदि परिकल्पना सत्य हो तो एक अच्छे परीक्षण के लिए उसे अस्वीकार करने की प्रायिकता बहुत कम होनी चाहिए। यदि वह सत्य न हो तो यह प्रायिकता बहुत अधिक होनी चाहिए।

§ १२'५'३ सिद्धान्त

इस तरह यदि दो परीक्षणों के लिए प्रथम प्रकार की त्रुटि की प्रायिकता बराबर हो जिसका परिमाण α हो तो इनमें से हम उस परीक्षण को चुनेंगे जिसके लिए असत्य परिकल्पना को अस्वीकार करने की प्रायिकता अधिक हो।

§ १२'६ परीक्षण सामर्थ्य और उसका महत्त्व

§ १२'६'१ परिभाषा—यदि परिकल्पना असत्य हो तो उसे अस्वीकार करने की प्रायिकता को परीक्षण-सामर्थ्य (power of test) कहते हैं।

§ १२'६'२ उदाहरण—हम सिद्धांत की भीमासा एक मामूली उदाहरण से आरंभ करेंगे। और इस उदाहरण की ही सहायता से कुछ नयी अवधारणाओं (concepts) को परिभाषा भी देंगे।

मान लीजिए कि प्रश्न है एक परिकल्पना के परीक्षण का जिसके अनुसार समष्टि का माध्य μ है। हम यह परीक्षण समष्टि पर बिना किसी प्रेक्षण के भी कर सकते हैं। कागज के छोटे-छोटे बिलकूल समान सौ टुकड़े कर लीजिए और उन पर क्रमशः एक से लेकर सौ तक की संख्याएँ लिख लीजिए। इन टुकड़ों को अच्छी-भाँति मिला लीजिए और इसके पश्चात् आँख बंद करके उनमें से एक को चुन लीजिए।

हमारा परीक्षण निम्नलिखित है—

यदि चुने हुए टुकड़े पर लिखी हुई संख्या 95 से अधिक हो तो परिकल्पना को अस्वीकार कर दीजिए, अन्यथा उसको स्वीकार कर लीजिए। क्योंकि इस परीक्षण

का उस समष्टि से कुछ स्रवव नहीं है जिसके स्रवव में परिकल्पना है, इसलिए यह मूर्खता-पूर्ण प्रतीत होता है, और है भी। परंतु यह ध्यान देने योग्य बात है कि यदि परिकल्पना सत्य है तो इस परीक्षण द्वारा उसके अस्वीकृत होने की प्रायिकता केवल 5% है। इस प्रकार इस परीक्षण के लिए $\alpha=0.05$ है और यदि परीक्षणों की तुलना करने के लिए हम केवल प्रथम प्रकार की त्रुटि का ही प्रयोग करते हैं तो यह परीक्षण उतना ही उत्तम है जितना कि प्रस्तुत समष्टि से चुने हुए एक हजार प्रेक्षणों पर आधारित ऐसा परीक्षण जिसके लिए भी $\alpha=0.05$ हो।

इनकी वास्तविक तुलना तो तब होनी है जब कि हम इन परीक्षणों की सामर्थ्य का पता लगाते हैं। मान लीजिए कि समष्टि का माध्य μ नहीं है बल्कि μ' है। हमारे कागज के टुकड़ोवाले परीक्षण द्वारा माध्य के μ होने की परिकल्पना के अस्वीकार किये जाने की प्रायिकता 5% है। इसलिए इस परीक्षण की सामर्थ्य $\beta=0.05$ है। यह एक ऐसा परीक्षण है जिसमें परिकल्पना के अस्वीकृत होने की प्रायिकता वही रहती है चाहे परिकल्पना सत्य हो और चाहे सत्य से बहुत दूर। यह स्थिति निश्चय ही अनवीपजनक है। परंतु इससे भी अधिक असवीपजनक स्थिति ही सकती है यदि सत्य होने पर भी परिकल्पना के अस्वीकृत होने की प्रायिकता α उसके असत्य होने पर अस्वीकृत होने की प्रायिकता β से भी अधिक हो। इस प्रकार की अवाञ्छनीय स्थिति उत्पन्न करने वाले परीक्षण को अभिनत परीक्षण (biased test) कहते हैं।

§ १२ ६ ३ अभिनत और अनभिनत परीक्षणों की परिभाषा

अभिनत परीक्षण—यह परीक्षण है जिसकी सामर्थ्य प्रथम प्रकार की त्रुटि की प्रायिकता से कम हो जाने $\beta < \alpha$ । जो परीक्षण अभिनत नहीं होता उसे अनभिनत (unbiased) कहते हैं।

§ १२ ७ प्राचल का अवकाश

क्योंकि हम यहाँ परिकल्पना परीक्षण के साधारण सिद्धान्तों की व्याख्या कर रहे हैं हमारे अध्ययन का क्षेत्र केवल माध्य अथवा प्रसरण से संबंधित परिकल्पनाओं तक ही सीमित नहीं रहना चाहिए। हम यहाँ समष्टि के किसी भी प्राचल से संबंधित परिकल्पना पर विचार करेंगे। यह प्राचल समष्टि के माध्यिका, चतुर्थी, तृतीय पूर्ण आदि में से कोई भी हो सकता है।

मान लीजिए कि हम Ω (ओमेगा) द्वारा प्राचल के अवकाश को सूचित करते हैं। इस अवकाश से हमारा तात्पर्य उन सब मानों के कुलक से है जो प्राचल के

लिए समभव हो। इस प्रकार प्रसामान्य वटन के माध्य के लिए— ∞ से लेकर $+\infty$ तक प्रत्येक मान धारण करना समभव है। इसलिए माध्य μ के लिए अवकाश Ω समस्त वास्तविक संख्याओं (real numbers) का कुलक है। प्रसामान्य वटन में ही प्रसरण σ^2 के लिए अवकाश केवल समस्त घनात्मक संख्याओं का कुलक है। द्विपद वटन में अनुपात P के लिए अवकाश 0 और 1 के बीच की संख्याएँ हैं।

§ १२.८ निराकरणीय परिकल्पना

मान लीजिए कि परिकल्पना यह है कि Ω के एक उपकुलक ω (ओमेगा का लघुरूप) में प्राचल θ (थीटा) स्थित है। इसको हम निम्नलिखित ढंग से सूचित करते हैं—

$$\theta \in \omega$$

और इसे θ स्थित है ω में पढ़ते हैं।

उदाहरण के लिए द्विपद वटन के अनुपात p के लिए परिकल्पना यह हो सकती है कि उसका मान 0.2 और 0.3 के बीच की कोई संख्या है। इस स्थिति में ω उन सब संख्याओं का कुलक है जो 0.2 और 0.3 के बीच में हैं। बहुधा इस उपकुलक ω में केवल एक ही संख्या होती है। उदाहरण के लिए इस परिकल्पना में कि समष्टि की माध्यिका 6 है, ω में केवल एक संख्या 6 ही है।

जिस परिकल्पना $\theta \in \omega$ का हम परीक्षण करते हैं उसे निराकरणीय परिकल्पना (null hypothesis) H_0 कहते हैं। वैकल्पिक परिकल्पना (alternative hypothesis) H_1 यह है कि '0 की स्थिति ω में नहीं है'। इसको हम निम्नलिखित संकेत से सूचित करते हैं

$$\theta \in (\Omega - \omega) = \omega'$$

यहाँ ω' अथवा $(\Omega - \omega)$ द्वारा हम Ω में स्थित उन राशियों को सूचित करते हैं जो ω में नहीं हैं।

§ १२.९ प्रतिदर्श और प्रतिदर्श-परिमाण

यह आवश्यक है कि परिकल्पना परीक्षण ऐसा होना चाहिए जो समष्टि पर किये हुए कुछ प्रेक्षणों पर आधारित हो। इन प्रेक्षणों के कुलक को प्रतिदर्श (sample) कहते हैं और प्रेक्षणों की संख्या को प्रतिदर्श-परिमाण (sample size)। यदि प्रतिदर्श

परिमाण μ हो और विभिन्न प्रेक्षण (x_1, x_2, \dots, x_n) हो तो हम इनके इस विशेष क्रम को \underline{x} से सूचित करते हैं।

$$(x_1, x_2, \dots, x_n) \equiv \underline{x} \quad (12.1)$$

§ १२.१० स्वीकृति और अस्वीकृति-क्षेत्र

\underline{x} के कुछ मान ऐसे होंगे जिनके लिए हम H_0 को अस्वीकार कर देंगे। इन सब मानों के कुलक C को परीक्षण का सशय-अंतराल (critical region) कहते हैं। इसी का दूसरा नाम अस्वीकृति-क्षेत्र भी है। \underline{x} के अन्य मानों के कुलक A को—जिनके लिए H_0 को अस्वीकार नहीं किया जाता—स्वीकृति-क्षेत्र (acceptance region) कहते हैं।

§ १२.११ प्रथम प्रकार की त्रुटि की प्रायिकता और सामर्थ्य

C पर आधारित परीक्षण के लिए प्रथम प्रकार की त्रुटि की प्रायिकता $\alpha(c)$ निराकरण्य परिकल्पना के सत्य होते हुए भी C में \underline{x} के पाये जाने की प्रायिकता है।

$$\alpha(c) = P[\underline{x} \in C \mid H_0] \quad (12.2)$$

किसी अन्य परिकल्पना H_1 के सत्य होने पर \underline{x} के C में पाये जाने की प्रायिकता को $\beta(c)$ से सूचित करते हैं और यह C पर आधारित परीक्षण का सामर्थ्य (power) है।

$$\beta(c) = P[\underline{x} \in C \mid H_1] \quad (12.3)$$

§ १२.१२ तुल्य तथा उत्तम परीक्षण

यदि C और C' दो अस्वीकृत क्षेत्र ऐसे हों जिनके लिए

$$\alpha(C) = \alpha(C')$$

$$\text{और } \beta(C) = \beta(C')$$

तो C और C' पर निम्न परीक्षणों को तुल्य (equivalent) समझा जाता है।

$$\text{यदि } \alpha(C) \leq \alpha(C')$$

$$\text{तथा } \beta(C) \leq \beta(C')$$

और यदि C और C' तुल्य न हों तो C को C' से उत्तम (superior) समझा जाता है।

§ १२.१३ प्रमेय

मान लीजिए H_0 के अनुसार x पर घनत्व फलन $f_0(x)$ है तथा H_1 के अनुसार $f_1(x)$ है और λ कोई घनात्मक संख्या है। यदि C_λ एक ऐसा अस्वीकृति-क्षेत्र है कि उसके किसी भी बिंदु x के लिए $f_1(x) \geq \lambda f_0(x)$ है तथा उसके बाहर किसी के भी बिंदु के लिए $f_1(x) < \lambda f_0(x)$ है, और C एक अन्य अस्वीकृति-क्षेत्र है तो इन अस्वीकृति-क्षेत्रों पर निर्भर परीक्षणों की सामर्थ्यों का अंतर इनकी प्रथम प्रकार की त्रुटियों की प्रायिकताओं के अंतर से कम-से-कम λ गुणा होगा।

उपपत्ति—

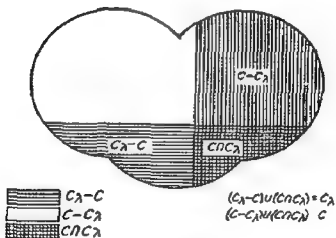
$$\begin{aligned} \text{सामर्थ्यों का अंतर} &= P[x \in C_\lambda | H_1] - P[x \in C | H_1] \\ &= P[x \in (C_\lambda - C) \cup (C \cap C_\lambda) | H_1] - P[x \in (C - C_\lambda) \cup (C \cap C_\lambda) | H_1] \\ &= \{P[x \in (C_\lambda - C) | H_1] + P[x \in (C \cap C_\lambda) | H_1]\} \\ &\quad - \{P[x \in (C - C_\lambda) | H_1] + P[x \in (C \cap C_\lambda) | H_1]\} \\ &= P[x \in (C_\lambda - C) | H_1] - P[x \in (C - C_\lambda) | H_1] \\ &\geq \lambda P[x \in (C_\lambda - C) | H_0] - \lambda P[x \in (C - C_\lambda) | H_0] \\ &= \lambda \{P[x \in C_\lambda | H_0] - P[x \in C | H_0]\} \\ &= \lambda [\alpha(C_\lambda) - \alpha(C)] \\ &= \lambda [\text{प्रथम प्रकार की त्रुटियों की प्रायिकताओं का अंतर}] \end{aligned}$$

यहाँ $(C - C_\lambda)$ से x के उन मानों के कुलक को सूचित किया गया है जो C में तो हैं परंतु C_λ में नहीं हैं। इसी प्रकार $(C_\lambda - C)$ से उन मानों के कुलक को सूचित किया गया है जो C_λ में हैं परंतु C में नहीं। चित्र संख्या 31 से यह अधिक स्पष्ट हो जायगा। इस उपपत्ति में इस ज्ञान का प्रयोग किया गया है कि

$$P[x \in (C_\lambda - C) | H_1] = \int_{(C_\lambda - C)} f_1(x) dx \quad \dots (12.4)$$

$$\begin{aligned} &\geq \lambda \int_{C_\lambda - c} f_0(x) dx \\ &= \lambda P[x \in (C_\lambda - C) | H_0] \end{aligned}$$

$$\begin{aligned} \text{और } P[x \in (C - C_\lambda) | H_1] &= \int_{(C - C_\lambda)} f_1(x) dx & (12.5) \\ &\leq \lambda \int_{(C - C_\lambda)} f_0(x) dx \\ &= \lambda P[x \in (C - C_\lambda) | H_0] \end{aligned}$$



चित्र ३१

§ १२.१४ ग्राह्य परीक्षण

यदि $\alpha(C_\lambda) = \alpha(C)$ तो हमें यह पता चलता है कि C_λ पर आधारित परीक्षण किसी भी ऐसे परीक्षण से कम सामर्थ्यवान् नहीं है जिसकी प्रथम प्रकार की भूल की प्रायिकता $\alpha(C_\lambda)$ है। इस प्रकार के परीक्षण को ग्राह्य (admissible) कहते हैं।

§ १२.१५ अस्वीकृति-क्षेत्र के चुनाव के अन्य निकष

नीमन पीयरसन सिद्धान्त के अनुसार हमें ऐसे परीक्षण को चुनना चाहिए जो ग्राह्य हो। ऊपर के प्रमेय द्वारा हम जानते हैं कि ग्राह्य परीक्षण को नैसर्गिक रूप से प्राप्त किया जा सकता है। हो सकता है कि आप परीक्षण के चुनाव के लिए किसी अन्य निकष को अधिक उत्तम समझे। उदाहरण के लिए आप वायद अस्वीकृति क्षेत्र को इस प्रकार चुनना अच्छा समझे कि दोनों प्रकार की त्रुटियों की प्रायिकता का कोई विशेष एक-धात फलन न्यूनतम हो जाय। आइए, देखें कि इस प्रकार के निकष के लिए अस्वीकृति क्षेत्र को ढूँढने का क्या तरीका हो सकता है।

यदि α_1 और α_2 द्वारा हम क्रमशः प्रथम और द्वितीय प्रकार की त्रुटियों की प्रायिकताओं को सूचित करें तो हमारा उद्देश्य एक ऐसे अस्वीकृति प्रदेश को मालूम करना है जो $p\alpha_1 + q\alpha_2$ को न्यूनतम कर दे जहाँ p और q दो घनात्मक ज्ञात सख्याएँ हैं।

किसी विशेष अस्वीकृति-क्षेत्र C के लिए

$$\begin{aligned} p\alpha_1(C) + q\alpha_2(C) &= pP[\underline{x} \in C \mid H_0] + qP[\underline{x} \in (\Omega - C) \mid H_1] \\ &= pP[\underline{x} \in C \mid H_0] + q(1 - P[\underline{x} \in C \mid H_1]) \\ &= q + \{pP[\underline{x} \in C, pf_0 > qf_1 \mid H_0] \\ &\quad - qP[\underline{x} \in C, pf_0 > qf_1 \mid H_1]\} \\ &\quad + \{pP[\underline{x} \in C, pf_0 < qf_1 \mid H_0] \\ &\quad - qP[\underline{x} \in C, pf_0 < qf_1 \mid H_1]\} \dots \dots \quad (12.6) \end{aligned}$$

यह स्पष्ट है कि प्रथम कुन्तल कोष्ठक (curled brackets) में ही हुई राशि घनात्मक तथा दूसरे कुन्तल कोष्ठक में ही हुई राशि ऋणात्मक है। इसलिए यदि कोई C के केवल उस भाग का अस्वीकृति-क्षेत्र की तरह उपयोग करता है जिसमें $pf_0 < qf_1$ हो तो इस नवीन अस्वीकृति क्षेत्र के लिए $p\alpha_1 + q\alpha_2$ का मान घट जायगा। इस प्रकार \underline{x} के जिन मानों के लिए $pf_0 < qf_1$ हो उन सबका कुलक सबसे उत्तम अस्वीकृति-क्षेत्र है, क्योंकि इसी में $p\alpha_1 + q\alpha_2$ न्यूनतम हो जाता है।

§ १२.१६ उदाहरण

निराकरणीय परिकल्पना H_0

X का बटन आयताकार (rectangular) है जिसका परास (0,2) है।

वैकल्पिक-परिकल्पना H_1

X का वटन आयताकार है जिसका परास $(1,5)$ है ।

$$\left. \begin{array}{l} \text{नहीं तो } f_0(x) = \frac{1}{4} \\ f_0(x) = 0 \end{array} \right\} \begin{array}{l} \text{यदि } 0 < x < 2 \\ \end{array} \dots\dots\dots(12.7)$$

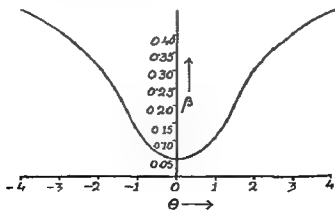
$$\left. \begin{array}{l} \text{नहीं तो } f_1(x) = \frac{1}{3} \\ f_1(x) = 0 \end{array} \right\} \begin{array}{l} \text{यदि } 1 < x < 5 \\ \end{array} \dots\dots\dots(12.8)$$

मान लीजिए कि हम उस अस्वीकृति-क्षेत्र को मालूम करना चाहते हैं जिसके लिए $2\alpha_1 + \alpha_2$ का मान न्यूनतम है । ऊपर के प्रमेय के अनुसार यह क्षेत्र ऐसा है जिसमें x के वे सब मान ऐसे हों जिनके लिए $2f_0(x) < f_1(x)$ हो और ऐसा कोई भी मान न हो जो इस असमता को सन्तुष्ट न करे ।

यह आसानी से देखा जा सकता है कि यह क्षेत्र $(2,5)$ है ।

§ १२.१७ कुछ परिभाषाएँ

§ १२.१७.१ सामर्थ्य-वक्र (power curve) परीक्षण की सामर्थ्य केवल अस्वीकृति-क्षेत्र पर ही नहीं बल्कि वैकल्पिक परिकल्पना पर भी निर्भर करती है । प्रत्येक



चित्र ३२— $\theta=0$ के एक परीक्षण का सामर्थ्य वक्र

मुनिश्चित वैकल्पिक परिकल्पना के लिए परीक्षण की एक विशेष सामर्थ्य होती है । इस सामर्थ्य को प्राचल का एक फलन समझा जा सकता है । प्राचल के विभिन्न मानों

के लिए यदि परीक्षण की सामर्थ्य को ग्राफ द्वारा चित्रित किया जाय तो एक वक्र प्राप्त होगा जो सामर्थ्य वक्र कहलाता है। चित्र ३२ में ऐसा सामर्थ्य वक्र दिखाया गया है जो निराकरणीय परिकल्पना $\theta = 0$ से संबंधित है।

§ १२.१७ २ एक-समान अधिकतम सामर्थ्यवान् परीक्षण (uniformly most powerful test)

यदि किसी परीक्षण की सामर्थ्य प्रत्येक विकल्प (alternative) के लिए किसी भी दूसरे परीक्षण की सामर्थ्य से अधिक हो तो उसे एक समान अधिकतम सामर्थ्यवान् कहा जाता है।

§ १२.१७.३ स्थानीयत अधिकतम सामर्थ्यवान् परीक्षण (locally most powerful test)

यदि निराकरणीय परिकल्पना से किसी विशेष विकल्प की तुलना करने के लिए एक परीक्षण दूसरे परीक्षणों की अपेक्षा अधिक सामर्थ्य रखता है, और यदि इसके लिए α का मान किसी अन्य परीक्षण के α से अधिक नहीं है तो उसे इस विकल्प के लिए स्थानीयतः अधिकतम सामर्थ्यवान् कहा जाता है।

§ १२.१७.४ एक-समान अनभिन्नत परीक्षण (Uniformly unbiased test)

यदि प्रत्येक विकल्प (प्राचल = θ) के लिए सामर्थ्य $\beta(\theta)$ प्रथम प्रकार की त्रुटि की प्रायिकता α से अधिक हो तो परीक्षण को एक-समान अनभिन्नत कहा जाता है।

§ १२.१७.५ स्थानीयत अभिन्नत परीक्षण (locally biased test)

यदि किसी विकल्प (प्राचल = θ_1) के लिए सामर्थ्य $\beta(\theta_1)$ प्रथम प्रकार की त्रुटि की प्रायिकता α से कम हो तो हम कहते हैं कि $\theta = \theta_1$ पर परीक्षण स्थानीयतः अभिन्नत है।

गणित द्वारा यह सिद्ध किया जा सकता है कि किसी भी विशेष विकल्प के लिए स्थानीयत अधिकतम सामर्थ्यवान् परीक्षण मालूम करना हमेशा संभव है और ये परीक्षण सदैव स्थानीयत अनभिन्नत भी होते हैं। इसके विपरीत यद्यपि कुछ विशेष परिकल्पनाओं के लिए एक समान अधिकतम सामर्थ्यवान् परीक्षण वर्तमान है, परंतु अन्य अनेक महत्वपूर्ण परिकल्पनाओं के लिए इस प्रकार का कोई परीक्षण संभव नहीं है। यदि किसी निराकरणीय परिकल्पना के लिए एक समान अधिकतम सामर्थ्यवान् परीक्षण वर्तमान है अथवा यदि हम उसके किसी विकल्प विशेष में ही

रखते हैं तो हमें उचित परीक्षण को चुनने में कुछ कठिनाई नहीं होगी। अन्यथा जो परीक्षण चुना जायगा उसका अन्य परीक्षणों से उत्तम होना प्राचल के वास्तविक मान पर निर्भर करेगा।

§ १२ १८ उदाहरण

एक प्रसामान्य समष्टि $N(\mu, \sigma)$ का प्रसरण σ^2 ज्ञात है और μ अज्ञात। इस समष्टि में से n परिमाण का एक यादृच्छिक प्रतिदर्श चुना जाता है। इसके आधार पर निराकरणिय परिकल्पना $\mu = \mu_0$ की परीक्षा करनी है।

यदि इन प्रेक्षणों को $\underline{x} = (x_1, x_2, \dots, x_n)$ से सूचित किया जाय तो इनका सयुक्त वटन निम्नलिखित होगा

$$f_1(\underline{x}) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \left[\left(\frac{x_1 - \mu}{\sigma} \right)^2 + \left(\frac{x_2 - \mu}{\sigma} \right)^2 + \dots + \left(\frac{x_n - \mu}{\sigma} \right)^2 \right]}$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2} \quad (12.9)$$

निराकरणिय परिकल्पना के अनुसार इनका सयुक्त वटन निम्नलिखित होगा।

$$f_0(\underline{x}) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2} \quad (12.10)$$

एक प्राह्य परीक्षण का पता चलाने के लिए हमें एक ऐसे अस्वीकृति क्षेत्र का पता चलाना है जिसके लिए

$$f_1(\underline{x}) \geq \lambda f_0(\underline{x})$$

$$\text{अथवा} \quad -\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2} \geq -\frac{\sum_{i=1}^n (x_i - \mu_0)^2}{2\sigma^2} + \log \lambda$$

$$\text{अथवा} \quad \frac{\sum_{i=1}^n (\mu - \mu_0)(2x_i - \mu - \mu_0)}{2\sigma^2} \geq \log \lambda$$

$$\text{अथवा} \quad \bar{x}(\mu - \mu_0) \geq \frac{1}{n} \left[\sigma^2 \log \lambda + \frac{\mu^2 - \mu_0^2}{2} \right]$$

यदि विकल्प यह हो कि $\mu > \mu_0$ तो अस्वीकृति क्षेत्र निम्नलिखित रूप से परिभाषित हो सकता है।

$$\bar{x} > k_1 \quad (12.11)$$

क्योंकि निराकरणाय परिकल्पना के आधार पर हमें \bar{x} का बटन ज्ञात है, इसलिए हम k_2 को इस प्रकार चुन सकते हैं कि \bar{x} के उससे अधिक होने की प्रायिकता एक पूर्व निश्चित सख्या हो। उदाहरण के लिए यदि यह सख्या ००५ हो तो हम जानते हैं कि $N(0, 1)$ का ५% बिन्दु १.९६ होता है

$$\therefore P \left[\frac{\bar{x} - \mu_0}{\sigma} > 1.96 \mid \mu = \mu_0 \right] = 0.05 \quad (12.12)$$

$$\text{इसलिए } k_2 = \mu_0 + 1.96 \sigma \quad (12.13)$$

इसी प्रकार यदि विकल्प $\mu < \mu_0$ हो तो अस्वीकृति-शेष की परिभाषा का निम्नलिखित रूप होगा।

$$\bar{x} < k_2 \quad (12.14)$$

इस प्रकार आप देखते हैं कि प्रसामान्य बटन में एक-तरफा विकल्पों के लिए जिस माध्य सवधी परीक्षण का साधारणतया उपयोग किया जाता है वह एक-समान अधिकतम सामर्थ्यवान् है।

§ १२.१९ नीमन-पीयरसन के सिद्धान्तों की आलोचना

इस विवेचना के बाद हम इस निश्चय पर पहुँचते हैं कि एक अच्छे परीक्षण के लिए प्राह्यता तथा अनभिन्नता के गुण आवश्यक हैं। यदि कोई परीक्षण एक-समान अधिकतम सामर्थ्यवान् हो तो निश्चय ही वह सर्वोत्तम है। परन्तु बहुत ही कम परिकल्पनाओं के लिए इस प्रकार के परीक्षण प्राप्त हैं। इनके अभाव में किसी अन्य निकष को अपनाया जाता है। ये अन्य निकष इतने तर्कपूर्ण और सतोपजनक नहीं हैं, और विभिन्न वैज्ञानिक विभिन्न निकषों को अधिक युक्ति-सयत मान सकते हैं।

यह भी संभव है कि विभिन्न अवस्थाओं में विभिन्न निकषों का प्रयोग उपयुक्त हो। प्रोफेसर रोटाल्ड ए० फिशर इसी कारण नीमन-पीयरसन के सिद्धान्त के कटु आलोचक हैं। उनका कहना है कि यद्यपि कुछ विशेष परिस्थितियों में, जहाँ वैकल्पिक परिकल्पनाओं को प्रस्तुत करना संभव है इन सिद्धांतों का प्रयोग किया जा सकता है, परन्तु साधारण वैज्ञानिक खोज में बहुधा हम विकल्पों से परिचित नहीं होते। ऐसी दशा में सामर्थ्य अथवा दूसरे प्रकार की त्रुटि की प्रायिकता पर विचार करना संभव नहीं है।

किसी ऐसी कहानी पर विश्वास करते हुए हम हिचकिचाहट का अनुभव करते हैं जिसके सच होने की संभावना बहुत कम हो। साधारणतया इस प्रकार की कहानी सुननेवालों पर निम्नलिखित प्रभाव पड़ सकते हैं—

- (१) यह सब अपोल-वल्पित है ।
- (२) ऐसा प्रतीत होता है कि घटना का वैज्ञानिक रीति से प्रेक्षण नहीं किया गया । घटना का वर्णन वास्तविकता से भिन्न है ।
- (३) कुछ बातें या तो बढ़ा-चढ़ा कर नहीं गयी हैं अथवा कुछ ऐसी घटनाओं का वर्णन नहीं किया गया है जो अवचित थीं और जिनसे इस कहानी की घटनाओं को समझने में सहायता मिलनी ।
- (४) कोई अन्य शक्ति अथवा कारण है जो हमारे वर्तमान ज्ञान की अवस्था में हमें अज्ञात है ।

इस प्रकार यदि किसी परिवर्तन को अस्वीकार किया जाता है तो यह आवश्यक नहीं है कि किसी विशेष वैकल्पिक परिवर्तन को स्वीकार किया जाय । और यदि हम किसी विशेष परिवर्तन को अस्वीकार नहीं करते तो इसका यह अर्थ नहीं है कि हम उस स्वीकार करते हैं । परिवर्तन को अस्वीकार करने में तर्क यह है कि अप्रामाणिक घटना के घटने पर विश्वास करने में हिचकिचाहट होती है । परंतु परिवर्तन को स्वीकार करने में इस प्रकार का कोई तर्क उपलब्ध नहीं होता ।

फिशर के अनुसार सारे सिद्धांत को इस पर आधारित करना अस्वीकृत-क्षेत्र चुनने का एक गलत दृष्टिकोण है कि यदि इस विरोध समष्टि पर इन्हीं परिस्थितियों में हजारों बार प्रयोग किया जाय तो केवल एक प्रतिशत अथवा पाँच प्रतिशत बार गलती होगी । कोई भी वैज्ञानिक एक ही सार्थकता-स्तर पर और एक ही समष्टि पर बार-बार प्रयोग नहीं करता । इसके अतिरिक्त प्रायिकता का परिवर्तन प्रायः ऐसी परिवर्तन पर आधारित होता है जिसकी संपूर्ण सत्यता पर किसी को विश्वास नहीं होता । उदाहरण के लिए जब हम इस परिवर्तन की जाँच करते हैं कि समष्टि प्रसामान्य है तो हम पहिले से ही जानते हैं कि यह यथार्थतः प्रसामान्य नहीं हो सकती । इस दशा में यदि हम दोनों प्रकार की त्रुटियों से बचना चाहते हैं तो सबसे सरल उपाय तो यह होता कि परिवर्तन को बिना परीक्षण के अस्वीकार कर दें । फिर भी हम परीक्षण करते हैं, क्योंकि हम वास्तव में यह जानना चाहते हैं कि प्रसामान्य घटना को समष्टि का प्रतिरूप (model) समझा जा सकता है या नहीं ।

§ १२२० फिशर की विचारधारा

फिशर चार प्रकार की परिस्थितियों में भेद करता है ।

§ १२२०१ बेज के प्रमेय का उपयोग

पहली परिस्थिति यह है जब समष्टि की पूर्वतः गृहीत प्रायिकताएँ (a-priori

probabilities) ज्ञात हो। हम इसके एक उदाहरण से पहिले ही परिचित हैं (देखिए § ३ १२)। हमें विभिन्न वर्तनों के चुनाव की प्रवृत्त गृहीत प्रायिकताएँ ज्ञात थी। चुनी हुई गोलियों के रंग के जानने पर हमें विभिन्न वर्तनों के चुनाव की प्रायिकताओं का परिकलन करना था। इस प्रकार की स्थिति में वेन के प्रमेय का उपयोग किया जाता है और प्रतिवधी प्रायिकता का परिकलन निम्नलिखित सूत्र से होता है—

$$P(A|B) = \frac{P(A) P(B|A)}{P(B)} \quad (12 15)$$

इस प्रकार हमें विभिन्न परिकल्पनाओं की प्रायिकताओं का ज्ञान होता है और यदि कोई निश्चय करना हो तो वह इन प्रायिकताओं के आधार पर किया जा सकता है। यदि किसी वैज्ञानिक को भविष्य में किये जानेवाले प्रयोगों के बारे में कुछ निश्चय करना है तो उसके लिए इन प्रायिकताओं का ज्ञान ही सफेद है यह घोषणा करने की कोई आवश्यकता नहीं है कि एक विशेष परिकल्पना सत्य है या असत्य।

परन्तु दुर्भाग्य से ऐसी परिस्थितियाँ बहुत कम होती हैं जब इस प्रकार के प्रायिकता सञ्चयी विवरण दिये जा सकते हों।

§ १२ २० २ प्रतिदर्श निरीक्षण योजना और नीमन-पीयरसन के सिद्धांतों का उपयोग

दूसरी परिस्थिति वह है जिसका औद्योगिक प्रक्रियाओं में बहुधा प्रादुर्भाव होता है। यदि प्रक्रिया नियंत्रित है तो उससे होनेवाले उत्पादन के लक्षणों का एक घटन होगा जिसे बहुत अधिक सख्या में प्रेक्षणों द्वारा जाना जा सकता है। यह उत्पादन कारखानों से बड़ी-बड़ी ढेरियों के रूप में निकलता है। समस्या यह जानना है कि किसी विशेष ढेरी में चूटिपूर्ण वस्तुओं की सख्या इतनी अधिक तो नहीं है कि इस प्रकार की ढेरी के बाजार में जाने से कारखाने के नाम पर घबरा लगने का डर हो। सिर्फ इस ज्ञान से ही काम नहीं चलेगा बल्कि इस प्रकार की ढेरियों को बाजार में जाने से रोकना पड़ेगा। इसके लिए ढेरियों का निरीक्षण करना होगा। परन्तु निरीक्षण में खर्च लगता है और यदि एक-एक वस्तु को परखा जाय तो वह महँगी हो जायगी—शायद इतनी महँगी कि उसकी खरीदने को कोई तैयार ही न हो। इस परिस्थिति में एक प्रतिदर्श-निरीक्षण योजना (sampling inspection plan) बनानी पड़ती है जिसमें चूटिपूर्ण उत्पादन के बाजार में जाने की सम्भावना कम हो जाय और खर्च भी अधिक न हो। इस दशा में निरीक्षक को ढेरी को बाजार में भेजने के लिए स्वीकृति या अस्वीकृति देना आवश्यक है और इस कार्य में दोनों प्रकार की चूटियाँ स्पष्ट ही हैं।

इसी प्रकार यह देखने के लिए कि उत्पादन नियंत्रण में है अथवा नहीं, उत्पादन होते समय ही बीच-बीच में से प्रतिदर्श चुने जा सकते हैं। प्रतिदर्श के आधार पर यह निर्णय करना होता है कि उत्पादन रोककर मशीन को ठीक करना चाहिए या नहीं। ऐसी स्थिति में जिस समष्टि के बारे में परिकल्पना का परीक्षण हो रहा है वह वास्तव में वर्तमान है और जिस प्राचल पर विचार किया जा रहा है उसका मान मालूम करना कठिन भले हो, परन्तु संभव है। इस प्रकार की समस्याओं को सुलझाने के लिए नीम्न-पीयरसन के सिद्धान्त विशेष उपयोगी हैं।

§ १२२०३ विश्वास्य युक्ति और पर्याप्त प्रतिदर्शन

सीसरी परिस्थिति वह है जो सबसे अधिक सामान्य है और वैज्ञानिक के लिए महत्वपूर्ण है। प्रायः परिकल्पना बहुत मुनिश्चित नहीं होती। कुछ प्राचलों के लिए किसी विनोप परास (range) के किसी भी मान को कारण बना इस परिकल्पना के अनुसार संभव होता है। उदाहरण के लिए जब हम कहते हैं कि समष्टि प्रसामान्य है तो हम कथन से समष्टि का पूरा विवरण नहीं मिलता। इस प्रसामान्य बंटन का $-\infty$ से $+\infty$ तक कोई भी माध्य हो सकता है और ० से $+\infty$ तक कोई भी प्रसरण हो सकता है। इस प्रकार की परिकल्पना के लिए आसजन सौष्टव (goodness of fit) के χ^2 -परीक्षण से आप पहिले ही परिचित हैं।

इस परीक्षण का पहला भाग होता है अज्ञात प्राचलों का प्राक्कलन करना। जब हमें इनके सर्वोत्तम प्राक्कलनों का ज्ञान हो जाता है तो इस ज्ञान और मूल परिकल्पना के संयोग से हमें समष्टि का एक पूरा विवरण प्राप्त हो जाता है। तब इस संपूर्ण विवरण की जांच की जाती है।

यद्यपि प्राक्कलन के सिद्धान्तों की विवेचना अभी तक नहीं की गयी, परन्तु यहाँ यह बताना आवश्यक है कि कुछ प्राक्कलक (estimators)* प्राचलों के बारे में वह सम्पूर्ण सूचना हमें दे देते हैं जो उनके आधारभूत आँकों से प्राप्त हो सकती है। ऐसे प्राक्कलक को पर्याप्त (sufficient) प्राक्कलक कहते हैं। यदि इस प्रकार का कोई प्राक्कलक विद्यमान हो तो एक नये प्रकार के तर्क का सहारा लिया जाता है जिसे विश्वास्ययुक्ति (fiducial argument) कहते हैं। इस युक्ति के प्रयोग

* प्रेक्षणे का वह फलन जिसके द्वारा किसी प्राचल का प्राक्कलन किया जाता है, उस प्राचल का प्राक्कलक कहलाता है।

पर एक और प्रतिबन्ध है। वह यह कि प्रेक्षण सावधानी से लिये हुए इस प्रकार के माप होने चाहिए कि उनको एक सतत चर के प्रेक्षित मान समझा जा सके और ऐसा समझने में कोई अर्थपूर्ण त्रुटि न हो।

मान लीजिए, प्राचल θ का इस प्रकार का एक प्राक्कलक $\hat{\theta}$ (थीटा-कलश) है। यदि हमें $\hat{\theta}$ का बटन ज्ञात है तो हम इस प्रकार की एक सख्या θ मालूम कर सकते हैं जिसके लिए $P[|\hat{\theta}-\theta| < \epsilon] = 0.95$

प्रेक्षणों के आधार पर θ का परिकलन किया जा सकता है और ऊपर के समीकरण में केवल θ ही अज्ञात है। इसलिए इस प्रायिकता-कथन (probability statement) को प्राचल का प्रायिकता-संबन्धी कथन समझा जा सकता है। इस प्रकार $\hat{\theta}$ के जानने से हमें θ का बटन मालूम हो सकता है। इस प्रायिकता बटन से यह निर्णय किया जा सकता है कि θ का एक बिन्दु परिकल्पित मान सम्भावी (probable) है अथवा नहीं। इस बटन के पाँच प्रतिशत अथवा एक प्रतिशत बिंदुओं का परिकलन किया जा सकता है। इनके आधार पर एक अस्वीकृत क्षेत्र का भी निर्माण किया जा सकता है।

किन्ती प्राचल को यादृच्छिक चर समझना कहाँ तक ठीक है यह विवादास्पद प्रश्न है। बंड के प्रमेय के सम्बन्ध में हम देख चुके हैं कि प्राचलों का भी एक पूर्वत गृहीत बटन हो सकता है। किसी विशेष समष्टि में जिसका अध्ययन किया जा रहा हो प्राचल का एक विशिष्ट मान होता है, परन्तु प्रेक्षण के पूर्व या तो यह प्राचल अज्ञात होता है या यादृच्छिक चर होता है। यदि हम जान जाते हैं कि प्राचल का मान क्या है तो यह यादृच्छिक चर नहीं बरन् एक ज्ञात अचर (constant) हो जाता है। इस प्रकार एक ही वस्तु यादृच्छिक चर अथवा अचर दोनों हो सकती हैं। वह क्या होगी यह इस पर निर्भर करता है कि उसके बारे में हमारा ज्ञान किस प्रकार का है।

यदि पूर्वत गृहीत बटन अज्ञात हो तो प्राचल की प्रतिष्ठा (status) भी एक अज्ञात (unknown) राशि की जैसी होती है। एक पर्याप्त प्रतिदर्श (sufficient statistic) के प्रेक्षण से प्राचल पूर्णतया ज्ञात तो नहीं होता, परन्तु नितान्त अज्ञात भी नहीं रहता, क्योंकि इस प्रतिदर्श से हमें प्राचल का कुछ तो ज्ञान ही हो जाता है। इस ज्ञान की प्रकृति प्रायिकता संबंधी होती है इसलिए प्राचल की प्रतिष्ठा अज्ञात से हटकर यादृच्छिक चर की हो जाती है।

§ १२ २०४ सभावितता फलन और उसका उपयोग

एक और परिस्थिति पर फिशर ने विचार किया है । यदि कोई पर्याप्त प्रतिदर्शज विद्यमान नहीं हो और प्राचल वा अवकाश असतत है तो ऊपर के तर्क से काम नहीं चल सकता । विभिन्न प्राचलको पर विचार करने से हमें प्राचल के विभिन्न घटन मिलेंगे और जब तक हमारे पास तर्क-भगत निक्षप (criterion) का अभाव है तब तक इनमें से किसी विशेष घटन का उपयोग करना और उसके आधार पर परिकल्पना को अस्वीकार करना असंगत होगा । इस अवस्था में फिशर के अनुसार हमें प्रायिकता को छोड़कर लगभग उसी के समान एक अन्य धारणा का आश्रय लेना होगा जिसे हम घटना की संभावितता (likelihood) की सजा देंगे ।

सभावितता प्राचल का एक फलन होता है । असतत घटनों के लिए इसका मान प्रेक्षित घटना की प्रायिकता के बराबर होता है । सतत घटनों के लिए—जहाँ किसी भी विशेष घटना की प्रायिकता धून्य होती है—इसका मान प्रेक्षित घटना के प्रायिकता—घनत्व के बराबर होता है । यह सभावितता प्राचल के किसी विशेष मान के लिए महत्तम होती है । जिन प्राचलों के लिए सभावितता फलन का मान महत्तम सभावितता की तुलना में बहुत कम होता है उन्हें सदैहजनक समझा जा सकता है ।

मान लीजिए, हम एक सिक्के को 25 बार उछालते हैं जिसमें वह 20 बार पट गिरता है । इस आधार पर हम इस परिकल्पना की जाँच करना चाहते हैं कि सिक्के के पट गिरने की प्रायिकता $\frac{1}{2}$ है । अभी तक हमने इसके जाँचने की जिस विधि पर विचार किया है उसमें हम परिकल्पना के आधार पर 20 या इससे भी अधिक बार पट आने की प्रायिकता का कलन करते हैं । यदि यह 25 प्रतिशत से कम हो तो हम परिकल्पना को अस्वीकार करते हैं (क्योंकि यहाँ हम दो-तरफा परीक्षण का उपयोग कर रहे हैं) । इस परीक्षण-प्रणाली की कई बार इस कारण आलोचना की गयी है कि तर्क और युक्ति में प्रेक्षित मान 20 को छोड़कर उससे भी बड़े अन्य मानों का उपयोग नहीं करना चाहिए ।

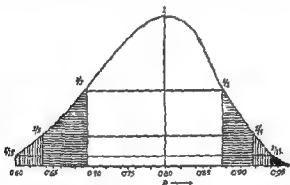
अच्छा यह होता कि प्रायिकता p के विभिन्न संभव मानों की तुलना प्रेक्षित बार-बारता के आधार पर की जाती । यदि पट गिरने की वास्तविक प्रायिकता p होती तो प्रेक्षित घटना की प्रायिकता, यदि त्रम को भी ध्यान में रखा जाता, $p^{20} (1-p)^5$ होती ।

इस उदाहरण में सभावितता $p = \frac{20}{25}$ के लिए महत्तम हो जाती है । p के

किसी भी मान के लिए इस सम्भाविता को महत्तम सम्भाविता के भिन्न (fraction) के रूप में रखा जा सकता है। इस उदाहरण में यह भिन्न निम्नलिखित है—

$$\left(\frac{p}{20/25}\right)^{20} \left(\frac{1-p}{5/25}\right)^5 = \left(\frac{p}{20}\right)^{20} \left(\frac{1-p}{5}\right)^5 (25)^{25}$$

हमें उस परिकल्पना को अस्वीकार करते हुए सबसे कम हिचकिचाहट होगी जिसके लिए सम्भाविता सबसे कम है और सबसे अधिक सम्भाविता वाली परिकल्पना को अस्वीकार करने में सबसे अधिक हिचकिचाहट होगी। यदि ऊपर के सम्भाविता-भिन्न तथा प्राचल के p सबष को दिखाते हुए एक ग्राफ खींचा जाय तो यह स्पष्ट हो जायगा कि प्राचल के ऐसे कौन-से मान हैं जिनकी सम्भाविताएँ महत्तम सम्भाविता से तुलना के योग्य हैं और किन् सीमाओं के बाहर सम्भाविता इतनी कम हो जाती है कि तत्सम्बन्धी प्राचल-मान सत्य-भासक (plausible) नहीं प्रतीत होते।



चित्र ३३— २५ में से २० बार सफलता के लिए p का सम्भाविता फलन

चित्र में p के परास को चार भागों में बाँटा गया है। (१) जहाँ सम्भाविता-भिन्न $\frac{1}{2}$ से अधिक है, (२) जहाँ यह $\frac{1}{2}$ से कम परन्तु $\frac{1}{3}$ से अधिक है, (३) जहाँ यह $\frac{1}{3}$ से कम परन्तु $\frac{1}{5}$ से अधिक है, और (४) जहाँ यह $\frac{1}{5}$ से भी कम है। अदिम भाग में प्राचल के मान स्पष्टतया सदेहजनक हैं। इस प्रकार p के परास को स्वीकृति और अस्वीकृति के क्षेत्रों में बाँटा जा सकता है। पर्याप्त प्राचलक (estimator) के अभाव में प्राचल के विभिन्न मानों की सत्यभासकता से परिचित कराने के लिए यह एक सग्त विधि है।

§ १२ २०५ वैज्ञानिक अध्ययन और स्वीकृति प्रणाली में अंतर

फिशर के मतानुसार वैज्ञानिक अध्ययन में परिकल्पना परीक्षण-अनुभव से सीखने और अपनी राय बदलने का एक साधन है । विज्ञान में राय कभी अंतिम नहीं होती तथा अधिक अनुभव होने पर वैज्ञानिक अपनी राय बदलने के लिए हमेशा स्वतंत्र रहता है । इस प्रकार परिकल्पनाओं को न तो सदा के लिए स्वीकार किया जाता है और न अस्वीकार । स्वीकृति प्रणाली (acceptance procedure) इस दृष्टिकोण से परिकल्पना-परीक्षण से भिन्न है । स्वीकृति प्रणाली में वर्तमान की एक विशेष समस्या पर कार्य करने के लिए निश्चय करना होता है जिसको बदलना संभव नहीं है । एक कारखाने के मालिक को यह तय करना पड़ता है कि वह किसी विशेष प्रकार का माल खरीदे अथवा नहीं । हो सकता है उसे बाद में अपनी गलती महसूस हो, परन्तु यह उस कच्चे माल को खरीदने और उसका उपयोग करने के बाद ही संभव है जिसके लिए स्वीकृति प्रणाली का प्रयोग किया जाता है । यह प्रणाली उस ही दशा में सगत है जब लगभग एक ही प्रकार के कच्चे माल पर बार-बार इसका प्रयोग किया जाय । इस प्रणाली में खर्च का विशेष ध्यान रखना पड़ता है । निरीक्षण का खर्चा उस जोखिम से अधिक नहीं होना चाहिए जो बिना निरीक्षण किये हुए माल को खरीदने में उठाया जाता है ।

परन्तु वैज्ञानिक गलत निर्णय से होनेवाले नुकसान को नहीं आँक सकता । वैज्ञानिकों की हैसियत से हम अनुमान लगाने की ऐसी पद्धति का उपयोग करना चाहते हैं जो सभी स्वतंत्र रूप से विचार करनेवालों के लिए युक्तिसंगत हो । इसका विचार हमारे सामने नहीं रहता कि इस अनुमान द्वारा प्राप्त ज्ञान का उपयोग किस प्रकार होगा ।

इस प्रकार आप देखते हैं कि नीमन-पीयरसन तथा फिशर के विचारों में भेद है और वास्तव में वे एक दूसरे के कट्टु आलोचक हैं । सीमाव्यवस्था विचारधारकों के भिन्न होते हुए भी कई समस्याओं के लिए दोनों के हल समान हैं । फिर भी हमेशा ऐसा नहीं होता कि जिस परिकल्पना को नीमन-पीयरसन के परीक्षण द्वारा अस्वीकार किया जाय वह संभावितता के उपयोग अथवा प्राचल के विश्वास्य-वटन (fiducial distribution) के प्रयोग से भी अस्वीकृत हो । आप इनमें से किसी के भी तर्कों से सहमत होने के लिए स्वतंत्र हैं, बल्कि यह भी हो सकता है कि आप को दोनों ही तर्कों में त्रुटि दृष्टिगोचर हो । अब हम परिकल्पना-परीक्षण के साधारण सिद्धांतों की विवेचना यही समाप्त करते हैं ।

भाग ३

सा ह च र्थ

समाश्रयण और सहसम्बन्ध

Association

Regression and Correlation

साहचर्य (Association)

§ १३१ परिचय

परिकल्पना-परीक्षण के सबध में हम कुछ ऐसे उदाहरणों से परिचय प्राप्त कर चुके हैं जिनमें प्रयोग का उद्देश्य यह जानना था कि दो विभिन्न गुणों में कोई सबध है या नहीं। इन परीक्षणों में समष्टि को एक $k \times r$ सारणी से विभाजित करके रखा जाता है जहाँ एक गुण के विचार से समष्टि के k भाग हैं और दूसरे गुण के विचार से r । इस सारणी में दोनों गुणों के स्वतंत्र होने की परिकल्पना के आधार पर विभिन्न खानों में प्रत्याशित बारबारता का परिकलन किया जाता है और χ^2 -परीक्षण द्वारा इन प्रत्याशित बारबारताओं और प्रेक्षित बारबारताओं के अन्तर की साध्यकता को आँका जाता है।

इस परीक्षण के अन्त में यदि χ^2 -का प्रेक्षित मान $\chi^2_{(k-1)(r-2)}$ के एक पूर्व निश्चित प्रतिशत बिंदु से अधिक हो तो हम निराकरणाय परिकल्पना का अस्वीकार कर देते हैं और इस निष्कर्ष पर पहुँचते हैं कि ये दो गुण स्वतंत्र नहीं हैं। अब प्रश्न यह उठता है कि यदि ये स्वतंत्र नहीं हैं तो इनके सबध को किस प्रकार समझा जा सकता है। यदि एक गुण में परिवर्तन होने पर दूसरे गुण में भी एक विशेष दिशा में परिवर्तन होने की प्रायिकता बढ़ जाय तो हम कहते हैं कि इन दोनों गुणों में साहचर्य (association) है।

§ १३२ साहचर्य की व्याख्या

गुणों में साहचर्य होने का क्या यह अर्थ है कि एक गुण दूसरे के साथ कारण और प्रभाव (cause and effect) के रूप में संबंधित है? जब हम औषध के सेवन और रोग से मुक्ति पाने में साहचर्य बताते हैं तो हमारा यही विचार होता है कि औषध के प्रभाव से रोगी अच्छे हो जाते हैं। यदि हम मशीनों की और उन पर धनी हुई वस्तुओं की त्रुटियों की समस्या में साहचर्य पाते हैं तो हमारा यही विश्वास होता

है कि अनुक मशीन अधिक अच्छी है और अमुक मशीन में कुछ दोष है। यदि मशीन में दोष न होता तो इतनी त्रुटियाँ उससे बनी हुई वस्तुओं में नहीं पायी जाती। हो सकता है कि हमारा इस प्रकार एक गुण को दूसरे का कारण समझना ठीक ही और यह भी हो सकता है कि यह हमारी भूल हो।

उदाहरण के लिए यदि हम यह देखने हैं कि किसी विशेष रोग में ऐलोपैथिक इलाज करवाने वाले रोगियों में नीरोग होने वालों का अनुपात अधिक है और वैद्यक इलाज करवाने वालों में कम, तो इसकी निम्नलिखित प्रकार की अनेक व्याख्याएँ की जा सकती हैं—

(१) इस रोग के लिए ऐलोपैथिक इलाज अधिक लाभदायक है।

(२) केवल संयोग से हमें ऐसे प्रेक्षण मिले हैं।

(३) ऐलोपैथिक इलाज करवाने वाले एक विशेष श्रेणी के लोग हैं जो वैद्यक इलाज करवाने वालों की अपेक्षा अधिक धनवान् हैं और इस इलाज के अतिरिक्त वे अधिक शक्तिवर्धक भोजन भी करते हैं। यही उनके स्वास्थ्य के रहस्य की कुजी है।

(४) रोग से मुक्ति पाने के लिए वैद्य अथवा डाक्टर पर विश्वास होना आवश्यक है। जिन लोगों ने वैद्यक इलाज करवाया उनमें से बहुतों को इस पर विश्वास न था। क्योंकि उनके पास ऐलोपैथिक इलाज के लिए पैसे नहीं थे इसलिए उन्हें मजदूरन वैद्यक का आश्रय लेना पड़ा। उनके स्वास्थ्य-लाभ न होने का कारण यह अविश्वास ही था।

ऐसी ही अन्य भी अनेक प्रकार की व्याख्याएँ प्रेक्षित सारणी के लिए की जा सकती हैं। परन्तु यह स्पष्ट है कि पहली व्याख्या के पक्ष में निर्णय देने से पहले हमें कम से कम तीसरी व्याख्या की जाँच अवश्य कर लेनी चाहिए।

इसी प्रकार यद्यपि विभिन्न मशीनों पर बनी वस्तुओं में त्रुटिसंख्या भिन्न-भिन्न हो सकती है, परन्तु इनका कारण मशीनों में अन्तर नहीं बरन् उन मजदूरों में अंतर हो सकता है जो इन पर काम करते हैं। इसी कारण प्रयोग की अभिकल्पना (design of experiments) के अध्याय में हम देखेंगे कि मशीनों में अन्तर के निष्कर्ष पर पहुँचने से पूर्व हमें अन्य कारणों के प्रभाव से मुक्ति पा लेना आवश्यक है। इसलिए लैटिन वर्ग (Latin Square) आदि अनेकों अभिकल्पनाओं (designs) का आविष्कार हुआ है। परन्तु कई स्थितियाँ ऐसी होती हैं जहाँ हम प्रयोग नहीं कर सकते, केवल समष्टि से एक प्रतिदर्श लेकर उस पर प्रेक्षण

सारणी सख्या 131

		पुत्रों की आँखों का रंग				
		काली (1)	भूरी (2)	नीली (3)	हरी (4)	कुल
पिता की आँखों का रंग	काली (1)	117	18	15	0	150
	भूरी (2)	55	180	15	0	250
	नीली (3)	0	12	60	3	75
	हरी (4)	0	0	1	24	25
	कुल	172	210	91	27	500

हम इस सारणी द्वारा पुत्रों की आँखों के रंग और उनके पिताजों की आँखों के रंग के साहचर्य का माप मालूम करना चाहते हैं। पुत्र से बिज्ञासा करने पर उसके पिता की आँख का रंग मालूम हो सकता है परन्तु पिता से पूछकर हम किसी होने वाले पुत्र की आँख का रंग नहीं मालूम कर सकते। लेकिन पिता की आँख के रंग के ज्ञान के आधार पर हम इसका अनुमान कर सकते हैं। पिता और पुत्र की आँखों के रंगों में जितना प्रगाढ़ साहचर्य होगा उतना ही अधिक हमें इस अनुमान पर विश्वास होगा। इस उदाहरण में साहचर्य के माप से हमारा उद्देश्य केवल यह जानना है कि पिता की आँख का रंग जानकर कितने विश्वास के साथ पुत्र की आँख के रंग के बारे में अनुमान लगाया जा सकता है।

यदि हम पिता की आँख का रंग जाने बिना यह अनुमान लगायें तो स्वाभाविक है कि हम वह रंग बतायेंगे जो सबसे अधिक पुत्रों में पाया जाता है। इस विशेष समष्टि के लिए यह रंग भूरा है। परन्तु कुल पुत्रों में केवल $\frac{210}{500} = 42\%$ की आँखों का यह रंग है इसलिए हमारे अनुमान के गलत होने की प्रायिकता 58% प्रतिशत है। प्रश्न उठता है कि पिता की आँख का रंग जानने से यह प्रायिकता कितनी कम हो जायगी।

पिता की आँख का रंग ज्ञात होने पर पुत्र की आँख के रंग का क्या अनुमान लगाना चाहिए? गलती की प्रायिकता को न्यूनतम करने के लिए यह स्वाभाविक है कि जिस

रंग की आँखवालों की संख्या उन सब पुत्रों में अधिकतम हो, जिनके पिता की आँख का वह ज्ञात-रंग है हम उसी रंग का अनुमान लगायें। जिन पुत्रों के पिता की आँख का रंग भूरा है उनमें सबसे अधिक संख्या भूरी आँखवालों की है। इसलिए यदि हमें यह पता हो कि पिता की आँख का रंग भूरा है तो हम पुत्र के बारे में भूरी आँख होने ही का अनुमान लगायेंगे। यह अनुमान $\frac{180}{250} = 72\%$ वार सत्य होगा। इसी नियम

के अनुसार पिता की आँख के रंग के आधार पर पुत्र की आँख के रंग का अनुमान करने से गलती की प्रायिकता नीली आँख के लिए $\frac{75-60}{75} = 20\%$ तथा काली आँख

के लिए $\frac{150-117}{150} = 22\%$ और हरी आँख के लिए केवल $\frac{25-24}{25} = 1\%$

है। यदि सब पुत्रों पर सम्मिलित विचार करें तो उन सब पुत्रों की संख्या जिनकी आँख के रंग का अनुमान पिता की आँख के रंग के आधार पर सही लगाया जायगा $117+180+60+24=381$ होगी। इस प्रकार गलती की कुल प्रायिकता $\frac{500-381}{500} = 23.8\%$ होगी।

ऊपर की तरह की सारणी में पक्ति के ज्ञान से स्तम्भ के अनुमान की गलती की प्रायिकता में जो आपेक्षिक कमी हो जाती है उसे g_{rc} से सूचित किया जाता है। इस उदाहरण की सारणी के लिए

$$g_{rc} = \frac{58.0 - 23.8}{58.0} \\ = 0.5896$$

इस तालिका में r से हम उस चर को सूचित करते हैं जिसके अनुसार पक्तियों (rows) का विभाजन किया गया है और c वह चर है जिसके अनुसार स्तम्भों (columns) को विभाजित किया गया है।

इसके विपरीत यदि हम पिता की आँख से पुत्र की आँख के रंग का अनुमान लगाने के स्थान पर पुत्र की आँखों के रंग से यह अनुमान लगायें कि पिता की आँख का रंग क्या रहा होगा तो इसमें स्तम्भ का स्थान प्रथम और पक्ति का स्थान द्वितीय होगा यानी स्तम्भ के दिव्ये हुए होने पर हम पक्ति का अनुमान लगायेंगे। इसके लिए उचित साहचर्य-सूचक (index of association) g_{cr} है।

$$g_{cr} = \frac{50.0 - 23.8}{50} \\ = 0.5240$$

लेकिन दोनों चरों में से एक के आधार पर दूसरे की प्रायिकता का कलन करने के बजाय हम दोनों के पारस्परिक साहचर्य के अनुमान के लिए ऐसे माप का कलन कर सकते हैं जो मूल में पिछले दोनों मापों के समान है परंतु उसका कलन ऐसे किया जाता है मानो आधे समय हम पक्ति को जान कर स्तंभ का अनुमान लगा रहे हो और आधे समय स्तंभ को जानते हुए पक्ति का। इस प्रकार की त्रुटि में जो कमी होगी वह पिछले दो मापों के अंशों (numerators) के योग को उनके हरों (denominators) के योग से विभाजित करने पर प्राप्त की जा सकती है। हम इस माप को g से सूचित करेंगे और इसे "पारस्परिक-साहचर्य" (mutual association) की उज्ञा देंगे। पिछली सारणी के आँकड़ों के अनुसार

$$\begin{aligned} g &= \frac{342 + 263}{580 + 500} \\ &= \frac{604}{1080} \\ &= 0.5593 \end{aligned}$$

मान लीजिए कि दो गुण शिक्षा और वेतन हैं। नीचे सरकारी कर्मचारियों को उनकी शिक्षा और वेतन के अनुसार एक क्रमबद्ध 5×4 सारणी में विभाजित किया हुआ है।

सारणी सख्या 132

सरकारी कर्मचारियों का शिक्षा और वेतन के क्रम के अनुसार वर्गीकरण

शिक्षा y	वेतन x				कुल
	$x < 100$	$100 \leq x < 300$	$300 \leq x < 500$	$500 \leq x$	
अपढ (1)	08	05	00	00	13
हाई-स्कूल (2)	11	14	03	00	28
इंटर मीडिएट (3)	12	23	04	00	39
ग्रजुएट (4)	07	104	35	16	162
पोस्ट ग्रेजुएट (5)	00	02	17	10	29
कुल	38	8	59	26	271

इस सारणी के लिए

$$g_{rc} = \frac{271 - 148}{271} - \frac{271 - (8 + 14 + 23 + 104 + 17)}{271}$$

$$\frac{271 - 148}{271}$$

$$= \frac{(8+14+23+104+17)-148}{(271-148)}$$

इसी प्रकार

$$g_{cr} = \frac{(12+104+35+16)-162}{(271-162)}$$

$$\begin{aligned} \therefore g &= \frac{((8+14+23+104+17)-148) + ((12+104+35+16)-162)}{(271-148) + (271-162)} \\ &= \frac{23}{232} \end{aligned}$$

§ १३.४ क्रमिक-साहचर्य का सूचकांक (index of order association)

इस माप g में एक कमी है। यदि वास्तविक तनस्वाह पाँच सौ रुपये से अधिक हो और हम यह अनुमान करें कि यह सौ रुपये से कम है अथवा यह अनुमान करें कि यह तीन सौ रुपये और पाँच सौ रुपये के बीच में है तो दोनों ही अनुमानों की त्रुटियों को इस माप में बराबर का दर्जा दिया गया है। इसी प्रकार इस माप में वेतन जानने पर हम शिक्षा के विचार से चाहे अथवा कर्मचारी के पोस्ट-ग्रेजुएट होने का अनुमान लगायें, चाहे उनके हाई स्कूल पास होने का—इन दोनों अनुमानों की त्रुटियों में भेद नहीं किया जाता। यदि दोनों चर इस प्रकार के हों कि उनको किसी तर्क-सयल क्रम में रखा जा सके जैसा कि उारपी सख्या 13 2 में है तो इन भूलों को बराबर समझना उचित नहीं प्रतीत होता। हम ऐसे माप की खोज करना चाहिए जो गूल की मात्रा से भी प्रभावित हो।

इस प्रकार का एक माप h है जिसे हम क्रमिक-साहचर्य का सूचकांक (index of order association) कहते हैं। यदि हम इन 271 कर्मचारियों में से दो को यादृच्छिकीकरण द्वारा चुन लें तो अधिक शिक्षाप्राप्त कर्मचारी के लिए अधिक वेतन होने की प्रायिकता कम वेतन होने की प्रायिकता से कितनी अधिक है? इस माप के लिए हम ऐसे चुनावों पर विचार नहीं करते जिनमें दोनों कर्मचारी वेतन अथवा शिक्षा के विचार से एक ही श्रेणी में रखे जा सकें।

§ १३.५ क्रमिक-साहचर्य के सूचकांक का कलन

इस माप को प्राप्त करने के निम्नलिखित विभिन्न चरण हैं

(1) हर एक खाने की बारबारता को उन सब बारबारताओं के योग से गुणा करिए जो उसके नीचे और दाहिने हाथ की ओर हो अर्थात् जिनमें X तथा Y दोनों का

मान अपेक्षाकृत बड़ा हो। उदाहरण के लिए पिछली सारणी में 23 का $(35+16+17+10) = 78$ से गुणा किया जायगा और 3 का $(16+10) = 26$ से। अंतिम पंक्ति और अंतिम स्तंभ की बारबारताओं को किसी भी संख्या से गुणा नहीं किया जाता।

(2) इन गुणनफलों का योग करिए। इस योग को यदि S से सूचित किया जाय तो सारणी के लिए

$$\begin{aligned} S &= (8 \times 228) + (5 \times 85) + (11 \times 211) \\ &\quad + (14 \times 82) + (3 \times 26) + (12 \times 184) \\ &\quad + (23 \times 78) + (4 \times 26) + (7 \times 29) \\ &\quad + (104 \times 27) + 35 \times 10 \\ &= 13,263 \end{aligned}$$

(3) प्रत्येक खाने की बारबारता को उन सब बारबारताओं से गुणा कीजिए जो उसके नीचे और बायीं ओर हैं अर्थात् जिनमें Y अपेक्षाकृत बड़ा हो किन्तु X अपेक्षाकृत छोटा हो।

(4) इस प्रकार के गुणनफलों का योग करके उसको D से सूचित करिए। पिछली सारणी में

$$\begin{aligned} D &= (5 \times 30) + (14 \times 19) + (23 \times 7) \\ &\quad + (3 \times 148) + (4 \times 113) + (35 \times 2) \\ &\quad + (16 \times 19) \\ &= 1,847 \end{aligned}$$

(5) h का परिकलन निम्नलिखित सूत्र से कीजिए

$$h = \frac{S-D}{S+D}$$

पिछली सारणी में

$$\begin{aligned} h &= \frac{13,263 - 1,847}{13,263 + 1,847} \\ &= \frac{11,416}{15,110} \end{aligned}$$

क्योंकि इस प्रकार के परिकलन में त्रुटि होने की संभावना है, इसलिए एक दूसरी प्रकार से इस परिकलन को करके दोनों परिकलनों के फल का मिलान किया जा सकता है। इसके लिए निम्नलिखित चरण हैं।

(ब) सब पक्ति-योगों और स्तम्भ-योगों के वर्गों के योग का परिकलन कीजिए और इसमें से खानों के वर्ग-योग को घटा दीजिए। यदि इस फल को n से सूचित किया जाय तो पिछली सारणी के लिए

$$\begin{aligned} n &= [(13)^2 + (28)^2 + (39)^2 + (162)^2 + (29)^2 \\ &\quad + (38)^2 + (148)^2 + (59)^2 + (26)^2] - [(8)^2 + (5)^2 + (11)^2 \\ &\quad + (14)^2 + (3)^2 + (12)^2 + (23)^2 + (4)^2 + (7)^2 + (104)^2 \\ &\quad + (35)^2 + (16)^2 + (2)^2 + (17)^2 + (10)^2] \\ &= 57,064 - 13,843 \\ &= 43,221 \end{aligned}$$

(ज) यदि परिकलन ठीक है तो

$$2(S+D) - n = n^2$$

जहाँ n सारणी के लिए कुल घाटघारता है।

इस सारणी में $n=273$ है।

$$\begin{aligned} \text{अब } 2(S+D) + n &= 30,220 + 43,221 \\ &= 73,441 \end{aligned}$$

और $n^2 = 73,441$

इसलिए हमने विश्वास है कि h का परिकलन सही हुआ है। h का मान -1 से लेकर $+1$ तक हो सकता है। यदि यह ऋणात्मक है तो इसका अर्थ यह है कि $D > S$ अर्थात् यदि कोई भी वो कर्मचारी लिये जायें और उनका क्रम दोनों चरों (x, y) के अनुसार मालूम किया जाय तो जिस कर्मचारी के लिए एक चर का मान अधिक होगा उसमें दूसरे चर का मान कम होने की आशा की जाती है। इसी प्रकार यदि h का मान धनात्मक है तो इसका अर्थ यह है कि $S > D$ अर्थात् जिस कर्मचारी के लिए एक चर का मान अधिक होगा उसमें दूसरे चर का मान भी अधिक होने की आशा की जाती है। यदि h का मान न्यूनतम अर्थात् -1 हो (जब $S=0$) अथवा अधिकतम यानी $+1$ हो (जब $D=0$) तो यह आशा निश्चय में परिणत हो जाती है।

§ १३.६ ऊपर के दिये हुए मापों का प्रयोग समष्टि और प्रतिदर्श दोनों के लिए किया जा सकता है। बहुधा समष्टि के लिए इस प्रकार का माप मालूम करना कठिन होता है और हम प्रतिदर्श से ही इस माप का प्राक्कलन (estimation) करते हैं।

ई जे वार हमारा यह विचार हो सकता है कि एक चर दूसरे से इस प्रकार संबंधित है जैसे कि कार्य और कारण। यदि कारण पर नियंत्रण रखा जाय तो कार्य भी नियंत्रित

हो सकता है। परन्तु प्रतिदर्श से प्राक्कलित माप के आधार पर इस निष्कर्ष पर पहुँचने में गलती की बहुत संभावना है। पहिले तो हमें यह विश्वास होना चाहिए कि प्रतिदर्श यादृच्छिकीकरण द्वारा चुना गया है। दूसरे यह ध्यान रखना चाहिए कि साहचर्य-सूचक का प्रेक्षित मान केवल प्रतिदर्श-त्रुटि के कारण तो संभव नहीं है। हमें यह भी पता होना चाहिए कि कोई तीसरा चर तो ऐसा नहीं है जो इन दोनों चरों को प्रभावित करता है। ऐसी दशा में इन दो चरों के साहचर्य का कारण यह तीसरा धर ही हो सकता है। ऐसे अनेक उदाहरण हैं जिनमें नौसिलिये सांख्यिक हास्यास्पद निष्कर्षों पर पहुँच जाते हैं क्योंकि वे ऊपर दी हुई बातों का ध्यान नहीं रखते। साहचर्य के मापों का परिकलन बहुत सरल है जिसे कोई भी स्कूल का विद्यार्थी सरलता से कर सकता है। परन्तु इस माप के आधार पर किसी युक्ति-युक्त निष्कर्ष पर पहुँचना बहुत सूझ-बूझ का काम है। यह सूझ-बूझ पुस्तकों द्वारा नहीं आ सकती वरन् केवल अनुभव और दूसरे सांख्यिकी की आलोचना से ही पायी जा सकती है।

अध्याय १४

सह-सम्बन्ध (Correlation)

§ १४ १ परिचय

x^2 परीक्षण और साहचय के संबंध में हम द्विचर (bivariate) से परिचय प्राप्त कर चुके हैं। साहचय के लिए हमने ऐसे चरों पर विचार किया था जिनको मापा नहीं जा सकता था—अधिक-से-अधिक किसी युक्ति-संगत क्रम में रखा जा सकता था। परन्तु आप जानते हैं कि कई चर ऐंसे होते हैं कि उनको मापा जा सकता है। इस प्रकार के चरों के बीच साहचय के लिए एक दूसरे ही प्रकार के माप का उपयोग किया जाता है। इस माप को सह-संबन्ध-गुणांक (Correlation coefficient) कहते हैं।

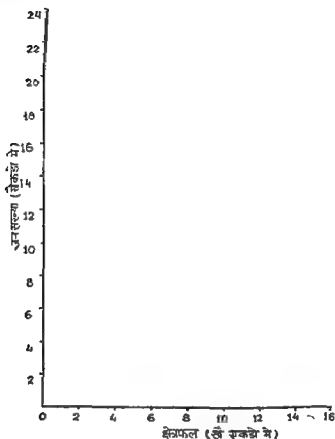
सारणी सख्या 14 1

ग्राम सख्या	क्षेत्र- फल	जन- सख्या	ग्राम सख्या	क्षेत्र- फल	जन- सख्या
i	x_i	y_i	i	x_i	y_i
(1)	(2)	(3)	(1)	(2)	(3)
1	3	8	9	5	10
2	4	5	10	5	1
3	6	10	11	10	7
4	5	5	12	8	3
5	11	6	13	4	2
6	15	20	14	4	10
7	15	10	15	6	6
8	11	5	16	4	6

§ १४ २ सह-संबन्ध सारणी

ऊपर की सारणी में सोलह गाँवों की जनसख्या सँकड़ों में और क्षेत्रफल सो एकड़ों में दिये हुए हैं। यह एक सह-संबन्ध सारणी का सबसे सरल उदाहरण है जिसमें प्रत्येक

इकाई के लिए दोनो चरो (x, y) के मान दिये हुए हैं। इन मानों को किसी विशेष क्रम में रखने की आवश्यकता नहीं है।



चित्र ३४—सारणी सख्या 14 I के लिए प्रकीर्ण-चित्र

§ १४३ घनात्मक व ऋणात्मक सहसंबंध

हम यह जानना चाहेंगे कि जब एक चर घटता या बढ़ता है तो दूसरा चर औसतन किस प्रकार विचलित होता है।

(I) यदि दोनो चरो X और Y के मान साथ-साथ बढ़ते हैं तो हम कहते हैं कि X और Y के बीच घनात्मक (positive) सहसंबंध है।

(2) यदि X के बढ़ने के साथ Y घटता है और X के घटने के साथ Y बढ़ता है तो हम कहते हैं कि X और Y का सह-सम्बन्ध ऋणात्मक (negative) है।

यह आवश्यक नहीं है कि जब X बढ़े तो Y या तो बढ़े ही या घटे ही। ऊपर की सारणी में X के बढ़ने पर कभी तो Y घटता है और कभी बढ़ता है। जब हम कहते हैं कि X और Y के बीच का सहसम्बन्ध ऋणात्मक है तो हमारा तात्पर्य केवल यह है कि साधारणतया X और Y साथ-साथ बढ़ते हैं।

इसके पहिले कि हम सहसम्बन्ध-गुणांक का परिकलन करें हमें कुछ साधारण सिद्धांतों का ध्यान रखना आवश्यक है। (1) यह निश्चय होना चाहिए कि इन दो चरों में कुछ सम्बन्ध होना न केवल सम्भव है बल्कि इस बात की आशा भी की जानी है। (2) यदि हमें यह नहीं मालूम कि कौन-सा गणितीय घटनसमष्टि का अच्छा प्रतिनिधित्व कर सकता है तो हमें केवल इस एक सख्या—सहसम्बन्ध गुणांक—से उतनी सूचना नहीं मिल सकती जितनी कि उस सारणी से जो इस परिकलन के लिए तैयार की जाती है।

(3) प्रगाढ़ सह-सम्बन्ध का अर्थ यह नहीं होता कि एक चर दूसरे के विचलन का कारण है।

§ १४४ प्रकीर्ण चित्र (Scatter diagram)

यदि हम एक ग्राफ पेपर में भुज (abscissa) पर x और कोटि (ordinate) पर y को सूचित करें तो x और y के प्रत्येक युग्म (pair) के लिए हमें एक बिंदु प्राप्त होगा। इस प्रकार सारणी अथवा न्यास (data) का लेखाचित्र पर बिंदुओं द्वारा निरूपण किया जा सकता है। इस तरह हमें जो चित्र प्राप्त होता है हम उसे प्रकीर्ण चित्र कहते हैं। उदाहरण के लिए सारणी सख्या 14 I के न्यास का प्रतिनिधित्व चित्र सख्या 34 में दिया हुआ है। इस चित्र के द्वारा हमें सहसम्बन्ध का माप नहीं मालूम हो सकता। यदि सारणी में दो या अधिक युग्म बिल्कुल समान हों तो उनकी बारबारताओं का हमें इस चित्र से पता नहीं चल सकता क्योंकि ये बिंदु संपतित हो जायेंगे और उनका पृथक् करना असम्भव होगा। न्यास द्वारा प्राप्त सूचना को प्रकीर्ण-चित्र में भूत्र रूप में रखने के लिए निम्नलिखित तरीका काम में लाया जाता है।

§ १४५ समाश्रयण-वक्र

X के प्रत्येक प्रेक्षित मान के लिए उससे सम्बन्धित Y के मानों के माध्य को इस प्रकीर्ण-चित्र पर एक बिंदु द्वारा सूचित किया जाता है। यदि न्यास एक बहुत बड़े प्रतिदर्श से लिया गया हो तो इन माध्य बिंदुओं को मिलानेवाली रेखा लगभग एक सतत

वक्र (smooth curve) होती है। इस वक्र को समाश्रयण-वक्र (regression curve) कहते हैं।

इसी प्रकार Y के हर प्रेषित मान के लिए X के माध्यों को मिलाने वाली रेखा एक दूसरा समाश्रयण-वक्र बनाती है। सबसे साधारण स्थिति में ये वक्र सरल रेखाएँ होते हैं और ऐसा समाश्रयण एक-घातक (linear) कहा जाता है। आगे हम अधिकतर एक-घातक समाश्रयण का ही अध्ययन करेंगे। ऊपर के प्रकीर्ण चित्र में इतने कम बिंदु हैं कि प्रत्येक X के मान के लिए Y का माध्य मालूम करना और एक सतत वक्र का पता चलाना व्यर्थ होगा। इसलिए केवल अनुमान से दो सरल रेखाएँ इस प्रकार खींची हुई हैं कि बिंदुओं से उनकी दूरी अधिक न हो।

इन दो समाश्रयण रेखाओं के खींचने के बाद समाश्रयण गुणांक का सन्निकट (approximate) मान मालूम किया जा सकता है। इस गुणांक का वास्तविक मान किस प्रकार परिकलित किया जाता है यह आगे बताया जायगा। परंतु इस वास्तविक मान का महत्त्व केवल उस समय है जब समाश्रयण एक-घातक अपना प्रायः एक घटक हो। प्रकीर्ण-चित्र द्वारा यह तय करने में बड़ी सहायता मिलती है कि समाश्रयण को एक घातक समझना कदा तक ठीक है।

§ १४६ सह-संबंध गुणांक (Correlation Coefficient)

यदि X और Y के माध्यों को हम क्रमशः \bar{X} और \bar{Y} से सूचित करें और X और Y में सहसंबंध घनात्मक हो तो हम यह आशा करते हैं कि यदि X का मान \bar{X} से कम होगा तो Y का मान भी \bar{Y} से कम होगा। इस प्रकार $(X - \bar{X})(Y - \bar{Y})$ का मान घनात्मक होगा। इसी प्रकार यदि X का मान \bar{X} से अधिक हो तो Y का मान भी \bar{Y} से अधिक होगा। इस दशा में भी $(X - \bar{X})(Y - \bar{Y})$ घनात्मक होगा। परंतु सहसंबंध के घनात्मक होने का यह अर्थ कदापि नहीं है कि प्रत्येक बिंदु के लिए $(X - \bar{X})(Y - \bar{Y})$ का मान घनात्मक ही होगा। इसका अर्थ केवल यह है कि औसतन इसका मान घनात्मक होना चाहिए।

अथवा

$$\frac{1}{N} \sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y}) > 0$$

इसी प्रकार जब सहसंबन्ध ऋणात्मक होता है तो

$$\frac{1}{N} \sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y}) < 0$$

यही नहीं बल्कि यदि सहसंबन्ध घनात्मक और प्रगाढ़ (strong) है तो

$$\frac{1}{N} \sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y}) \text{ का मान घनात्मक और बड़ा होता है। यदि सह-}$$

संबन्ध घनात्मक तो हो, परन्तु निर्बल (weak) हो तो यह मान घनात्मक और अपेक्षाकृत छोटा होता है। इसी प्रकार ऋणात्मक सहसंबन्ध प्रगाढ़ अथवा निर्बल होने के अनु-

सार $\frac{1}{N} \sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y})$ का मान ऋणात्मक और क्रमशः छोटा अथवा बड़ा होता है।

इससे यह प्रतीत होता है कि कदाचित् $(X - \bar{X})(Y - \bar{Y})$ का प्रत्याशित मान

$$C_{xy} = E(X - \bar{X})(Y - \bar{Y})$$

सहसंबन्ध का एक अच्छा माप है। परन्तु इसका मान उन मात्रको (units) पर निर्भर करता है जिनमें X और Y को मापा जाय। क्योंकि सहसंबन्ध दो गुणों के संबन्ध का माप है, इसलिए हम यह चाहेंगे कि वह इन गुणों के मात्रको से स्वतंत्र हो। उदाहरण के लिए यदि हम यह जानना चाहें कि गाँवों के शस्य-क्षेत्रफल और संपूर्ण क्षेत्रफल में संबन्ध अधिक प्रगाढ़ है अथवा शस्य-क्षेत्रफल और किसानों की संख्या में, तो C_{xy} की तरह का माप हमारे काम में नहीं आ सकता।

यदि X को उसके बटन के मानक विचलन σ_x के मात्रको से और Y को उसके बटन के मानक विचलन σ_y के मात्रको से मापा जाय तो यह समस्या हल हो जायगी, क्योंकि इस दशा में C_{xy} केवल एक संख्या होगी जिसमें कोई मात्रक समाविष्ट नहीं है।

X और Y को σ_x और σ_y के मात्रको में मापने का अर्थ है कि X के स्थान पर $\frac{X}{\sigma_x}$ तथा

Y के स्थान पर $\frac{Y}{\sigma_y}$ का उपयोग करना। इस प्रकार से प्राप्त C_{xy} के मान को हम

r से सूचित करेंगे।

$$\begin{aligned}
 r &= \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i}{\sigma_x} - \frac{\bar{x}}{\sigma_x} \right) \left(\frac{y_i}{\sigma_y} - \frac{\bar{y}}{\sigma_y} \right) \\
 &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{X}) (y_i - \bar{Y}) \\
 &\quad \frac{\sigma_x \sigma_y}{\sigma_x \sigma_y} \\
 &= \frac{C_{xy}}{\sigma_x \sigma_y}
 \end{aligned} \tag{14.1}$$

इस नये माप r को जो मात्रको से स्वतंत्र है सहसंबंध गुणांक (correlation coefficient) कहते हैं।

§ १४.७ समाश्रयण गुणांक और सहसंबंध गुणांक में संबंध

हम समाश्रयण रेखाओं का पहिले ही बणन कर चुके हैं। हम देखेंगे कि इन रेखाओं के समीकरण निम्नलिखित हैं।

$$\frac{Y - \bar{Y}}{\sigma_y} = \frac{C_{xy}}{\sigma_x \sigma_y} \frac{X - \bar{X}}{\sigma_x}$$

$$\text{अथवा } (Y - \bar{Y}) = r \frac{\sigma_y}{\sigma_x} (X - \bar{X}) \tag{14.2}$$

$$\text{तथा } (X - \bar{X}) = r \frac{\sigma_x}{\sigma_y} (Y - \bar{Y}) \tag{14.3}$$

ये दोनो समीकरण क्रमशः Y के X पर तथा X के Y पर समाश्रयण को सूचित करते हैं। $r \frac{\sigma_y}{\sigma_x}$ तथा $r \frac{\sigma_x}{\sigma_y}$ को समाश्रयण गुणांक (regression coefficients) की संज्ञा दी जाती है।

इस प्रकार

$$b_{yx} = r \frac{\sigma_y}{\sigma_x} = Y \text{ का } X \text{ पर समाश्रयण-गुणांक}$$

$$b_{xy} = r \frac{\sigma_x}{\sigma_y} = X \text{ का } Y \text{ पर समाश्रयण गुणांक}$$

$$\therefore b_{yx} \cdot b_{xy} = \frac{r \sigma_y}{\sigma_x} \cdot \frac{r \sigma_x}{\sigma_y} = r^2 \quad \dots\dots\dots(14.4)$$

§ १४.८ सह-संबंध-गुणांक का परिकलन

r का मान प्राप्त करने के लिए \bar{X} , \bar{Y} , σ_x , σ_y और C_{xy} का परिकलन आवश्यक है। आप \bar{X} , \bar{Y} , σ_x और σ_y के परिकलन से तो पहिले ही परिचित हैं। C_{xy} के परिकलन के लिए भी एक सरल तरीका है।

$$C_{xy} = \frac{1}{N} \sum_{j=1}^N (x_j - \bar{X})(y_j - \bar{Y}) = \frac{1}{N} \sum_{j=1}^N xy_j - \bar{X} \bar{Y} \quad \dots\dots\dots(14.5)$$

$$\therefore r = \frac{\frac{1}{N} \sum_{j=1}^N xy_j - \bar{X} \bar{Y}}{\sqrt{\left[\frac{1}{N} \sum_{j=1}^N x_j^2 - \bar{X}^2 \right] \left[\frac{1}{N} \sum_{j=1}^N y_j^2 - \bar{Y}^2 \right]}} = \frac{\sum_{j=1}^N x_j y_j - \bar{Y} \sum_{j=1}^N x_j}{\sqrt{\left[\sum_{j=1}^N x_j^2 - \bar{X} \sum_{j=1}^N x_j \right] \left[\sum_{j=1}^N y_j^2 - \bar{Y} \sum_{j=1}^N y_j \right]}} \quad \dots\dots\dots(14.6)$$

सारणी सत्या 14.1 के लिए r का परिकलन नीचे दिया हुआ है।

N=16	$\sum_{j=1}^{16} x_j = 116$	$\sum_{j=1}^{16} y_j = 116$
------	-----------------------------	-----------------------------

$$\therefore \bar{X} = \frac{116}{16} = 7.25 \quad \therefore \bar{Y} = 7.25$$

$$\sum_{i=1}^{16} x_i^2 = 1,076 \quad \therefore \sum_{i=1}^{16} x_i^2 - \bar{X} \sum_{i=1}^{16} x_i = 235$$

$$\sum_{i=1}^{16} y_i^2 = 1,142 \quad \therefore \sum_{i=1}^{16} y_i^2 - \bar{Y} \sum_{i=1}^{16} y_i = 301$$

$$\sum_{i=1}^{16} x_i y_i = 977 \quad \therefore \sum_{i=1}^{16} x_i y_i - \bar{Y} \sum_{i=1}^{16} x_i = 136$$

$$\begin{aligned} \therefore r &= \frac{136}{\sqrt{235 \times 301}} \\ &= \frac{136}{265.94} \\ &= 0.5114 \end{aligned}$$

§ १४.९ बहुत बड़े प्रतिदर्श के लिए सहसंबंध गुणांक का परिकलन

यदि कुल प्रतिदर्श में केवल 25 या 30 प्रेक्षण हों तो इस प्रकार सह-संबंध गुणांक का परिकलन करने में अधिक कठिनाई नहीं होती। परंतु यदि प्रतिदर्श बड़ा हो, उसमें सैंकड़ों अथवा हजारों प्रेक्षण हों तो इस प्रकार परिकलन संभव होते हुए भी कठिन है और इसमें त्रुटि होने की संभावना बहुत अधिक हो जाती है। जिस प्रकार हम घर के परास (range) को कुछ अंतरालों में विभाजित करके—और यह मानकर कि अंतरालों के सभी प्रेक्षण उसके मध्य बिंदु पर स्थित है—प्रसरण के परिकलन को सरल बना लेते हैं, उसी प्रकार हम सह-संबंध गुणांक के परिकलन को भी सरल बना सकते हैं। इस तरीके को नीचे के उदाहरण द्वारा समझाने की चेष्टा की गयी है।

194 खेतों में प्रति एकड़ उपज Y (बुशलों में) और उनमें डाले हुए नाइट्रोजन खाद का परिमाण X (पाउण्डों में) सारणी 14.2 में दिये हुए हैं। हम इन आँकड़ों के आधार पर उपज और खाद के परिमाण के सह-संबंध गुणांक का परिकलन करेंगे। इन परिकलनों के कई चरण इस सारणी के साथ ही दिये हुए हैं।

१४.९१ परिकलन की जाँच

क्योंकि इतने लंबे परिकलन में गलती हो जाने की संभावना है, इसलिए हर एक परिकलन की जाँच करना आवश्यक है। यह देखा गया है कि यदि एक ही परिकलन

सारणी संख्या 14.2

नास्ट्रोमन खाद का परिष्कार

Y	X	0-20	20-40	40-60	60-80	80-100	100-120	120-140	140-160	f_y	Yf_y	$\sum Yf_y$	Y^2f_y	$\sum Y^2f_y$	$Y^2 \sum f_y$	$\sum Y^2 f_y$
0-4	$\frac{x}{2}$	-3	-2	-1	0	1	2	3	4	6	-30	-18	90	54	150	90
4-8	$\frac{x}{2}$	6								10	-40	-30	120	90	160	120
8-12	$\frac{x}{2}$	4	4							8	-24	-20	60	52	72	60
12-16	$\frac{x}{2}$		8							8	-16	-16	32	32	32	32
16-20	$\frac{x}{2}$		30	2						12	-12	-22	22	42	12	22
20-24	$\frac{x}{2}$		2	12						14	0	-16	0	20	0	0
24-28	$\frac{x}{2}$			15	20	3				39	39	-14	22	118	39	-14
28-32	$\frac{x}{2}$			1	18	1	9	5	1	54	108	56	216	118	216	118
32-36	$\frac{x}{2}$				6		10	4	8	43	139	79	387	219	387	237
	f_x	20	25	30	44	38	19	9	9	194	154	-1	1068	649	1068	649
	$\sum f_x$	-60	-50	-30	0	38	38	27	36	-1						
	$\sum Y^2 f_y$	-82	-37	15	74	88	48	22	26	154						
	$\sum Y^2 f_x$	180	100	30	0	38	76	81	144	639						
	$\sum Y^2 f_{xy}$	346	79	21	146	218	126	56	76	1060						
	$\sum Y^2 f_{xy}$	246	74	-15	0	33	96	66	104	639						

को एक ही मनुष्य द्वारा करता है तो गलती के दुहराये जाने की काफी संभावना रहती है। इसलिए यदि हो सके तो परिकलन को जाँचने के लिए किसी दूसरी विधि का प्रयोग करना चाहिए। इस सारणी में प्रत्येक परिकलन को दो प्रकार से किया गया है। यदि इन दोनों में अंतर हो तो अधिक बारीकी से निरीक्षण करके भूल का पता चलाया जा सकता है।

उपर्युक्त सारणी में किसी विशेष (x, y) खाने की बारंबारता को f_{xy} से सूचित किया गया है। इसी प्रकार किसी विशेष x' अंतराल की बारंबारता को $f_{x'}$ तथा किसी विशेष y' अंतराल की बारंबारता को $f_{y'}$ से सूचित किया गया है।

§ १४.१० मूलबिंदु व मात्रक का परिवर्तन

परिकलन की सरलता के लिए मूल बिंदु (origin) तथा मात्रको (units) को बदल दिया गया है। इस विधि से अध्याय २ में, प्रसरण के कलन के राशियों में, आप पहिले ही परिचित हो चुके हैं।

इस सारणी में

$$N = \sum f_{xy} = \sum f_{x'} = 194 ; \sum_{i=1}^{194} x'_i = \sum_{x'} x' f_{x'} = \sum_{x'} \sum_{y'} x' f_{xy} = -1$$

$$\sum_{i=1}^{194} x'^2_i = \sum_{x'} x'^2 f_{x'} = \sum_{x'} \sum_{y'} x'^2 f_{xy} = 649$$

$$\sum_{i=1}^{194} y'_i = \sum_{y'} \sum_{x'} y' f_{xy} = \sum_{y'} y' f_{y'} = 154$$

$$\sum_{i=1}^{194} y'^2_i = \sum_{y'} \sum_{x'} y'^2 f_{xy} = \sum_{y'} y'^2 f_{y'} = 1,068$$

$$\sum_{i=1}^{194} x'_i y'_i = \sum_{x'} x' \sum_{y'} y' f_{xy} = \sum_{y'} y' \sum_{x'} x' f_{xy} = -659$$

$$\therefore r = \frac{659 - \frac{194 \times (-1)}{194}}{\sqrt{\left(649 - \frac{1}{194}\right) \left(1068 - \frac{(154)^2}{194}\right)}}$$

$$\begin{aligned}
 &= \frac{659\ 7938}{\sqrt{648.9949 \times 945\ 7526}} \\
 &= \frac{659\ 7938}{783.4466} \\
 &= 0.8422
 \end{aligned}$$

यह ध्यान देने योग्य बात है कि मूलबिंदु और मात्रको के बदलने से r के मान पर कोई प्रभाव नहीं पड़ता क्योंकि (1) $x_i - \bar{X}$ तथा $y_i - \bar{Y}$ क्रमशः X और Y के बदलने के माध्यों से x'_i और y'_i के अंतर हैं और ये मूलबिंदु पर निर्भर नहीं करते। (2) यदि x_i और y_i को किन्हीं अचल राशियों C_1 और C_2 से गुणा किया जाय और गुणनफल को x'_i और y'_i से सूचित किया जाय तो

$$\sum_{i=1}^N (x'_i - \bar{X}') (y'_i - \bar{Y}') = C_1 C_2 \sum_{i=1}^N (x_i - \bar{X}) (y_i - \bar{Y})$$

$$\sum_{i=1}^N (x'_i - \bar{X}')^2 = C_1^2 \sum_{i=1}^N (x_i - \bar{X})^2$$

$$\sum_{i=1}^N (y'_i - \bar{Y}')^2 = C_2^2 \sum_{i=1}^N (y_i - \bar{Y})^2$$

∴ $x' = \underline{X} C_1$ और $y' = \underline{Y} C_2$ का सहसंबंध गुणांक यदि $r'_{x'y'}$ हो तो

$$\begin{aligned}
 r'_{x'y'} &= \frac{\sum_{i=1}^N (x'_i - \bar{X}') (y'_i - \bar{Y}')}{\sqrt{\left[\sum_{i=1}^N (x'_i - \bar{X}')^2 \right] \left[\sum_{i=1}^N (y'_i - \bar{Y}')^2 \right]}} \\
 &= C_1 C_2 \frac{\sum_{i=1}^N (x_i - \bar{X}) (y_i - \bar{Y})}{\sqrt{\left[C_1^2 \sum_{i=1}^N (x_i - \bar{X})^2 \right] \left[C_2^2 \sum_{i=1}^N (y_i - \bar{Y})^2 \right]}}
 \end{aligned}$$

$$\begin{aligned} &= \frac{N}{\sum_{i=1}^N} (x_i - \bar{X})(y_i - \bar{Y}) \\ &= \frac{\sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\left[\sum_{i=1}^N (x_i - \bar{X})^2 \right] \left[\sum_{i=1}^N (y_i - \bar{Y})^2 \right]}} \\ &= r \end{aligned}$$

घक-आसंजन (Curve Fitting)

§ १५.१ अनुमान में त्रुटि

मान लीजिए कि (X, Y) एक द्विचर है। इसमें हमें X का मान ज्ञात है और हम Y के मान का अनुमान लगाना चाहते हैं। यह स्पष्ट है कि हम Y के केवल उन मानों पर विचार करेंगे जो X के इस मान के साथ सम्भव हैं। मान लीजिए कि Y के ये मान $y_1, y_2, y_3, \dots, y_{n-1}, y_n$ हैं। अब यदि हमारा अनुमान $Y = y'$ हो तो इसमें कुछ त्रुटि हो सकती है। यदि वास्तविक मान y_1 हो तो यह त्रुटि $(y' - y_1)$ है। और यदि वास्तविक मान y_2 हो तो यह त्रुटि $(y' - y_2)$ है। इसी प्रकार Y के विभिन्न मानों के लिए विभिन्न त्रुटियाँ होंगी। आपको शायद आश्चर्य हो कि आखिर यह अनुमान लगाने का प्रश्न क्यों उठता है। स्थिति स्पष्ट करने के लिए हम एक उदाहरण पर विचार करेंगे।

हमें मालूम है कि किसी परिवार की आय बढ़ने के साथ कपडों पर उसका खर्चा भी बढ़ता है। यह इतना स्पष्ट है कि दोनों चरों की स्वतंत्रता की परिकल्पना की जाँच करना अनावश्यक है। इनके सह-संबंध गुणांक का मान मालूम करने से भी कुछ विशेष लाभ प्रतीत नहीं होता। देश के लिए योजना बनाने वाले यह जानना चाहेंगे कि परिवार की आय जानने पर क्या कपडों पर उसके खर्च का अनुमान लगा सकते हैं। इस प्रकार यदि उन्हें देश में आय का वितरण ज्ञात हो तो उन्हें यह पता चल सकता है कि देश के लिए कुल कितने कपडे की आवश्यकता होगी।

ये अनुमान त्रुटिपूर्ण हो सकते हैं। एक ही आयवाले अनेक परिवार हो सकते हैं, परन्तु उन सबका कपडों पर खर्च बराबर नहीं होगा। यदि हम इनमें से किसी एक i -वें परिवार के कपडे पर खर्च का अनुमान y' लगायें और वास्तविक खर्च y_i हो तो त्रुटि $(y' - y_i)$ होगी। क्योंकि यह अनुमान केवल आय X पर निर्भर करता है, इसलिए उन सभी परिवारों के लिए जिनकी आय X है खर्च का अनुमान y' ही होगा और त्रुटियाँ क्रमशः $(y' - y_1), (y' - y_2), \dots, (y' - y_n)$ होगी।

अब प्रश्न यह है कि खर्च का अनुमान किस प्रकार लगाया जाय । इसके लिए हम ऐसे परिवारों का एक यादृच्छिक प्रतिदर्श ले सकते हैं जिनकी आय x हो । इनके रूपड़ों के खर्च के प्रेक्षित मानों के आधार पर हम ऐसे मान y' को निर्धारित कर सकते हैं जिससे इन प्रेक्षित मानों का औसत अंतर न्यूनतम हो । यदि प्रतिदर्श समष्टि वा एक अच्छा प्रतिनिधि हो तो इस y' को x आयवाले परिवारों के लिए कपड़े पर खर्च के प्रतिनिधि रूप में रखा जा सकता है । यह आप पहिले से ही जानते हैं कि यदि इस प्रतिनिधि को y के प्रेक्षित मानों का माध्य लिया जाय तो त्रुटियों का वर्ग-योग न्यूनतम होगा ।

$$\begin{aligned} \therefore \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n \{(y_i - a) - (\bar{y} - a)\}^2 \\ &= \sum_{i=1}^n (y_i - a)^2 - n(\bar{y} - a)^2 \\ &\leq \sum_{i=1}^n (y_i - a)^2 \end{aligned}$$

जहाँ कोई भी अन्य कल्पित प्रतिनिधि है ।

परतु योजना बनाने वाली को कितनी विशेष आय x में ही विशेष बिलचस्पी नहीं है । वे ही x के प्रत्येक मान के लिए y का अनुमान जानना चाहेंगे । यदि x के प्रत्येक मान के लिए परिवारों का अलग-अलग प्रतिदर्श लिया जाय तो कुल प्रतिदर्श बहुत बड़ा हो जायगा । इसके अतिरिक्त साधारणतया हमारे पास परिवारों की ऐसी सूची नहीं होती जिसमें उनकी आय भी दी हुई हो । परिवारों को चुनने और उनसे प्रश्न करने पर ही हमें मालूम हो सक्ता है कि उनकी आय क्या है । प्रत्येक विशेष आय के अनेक परिवार चुनने के लिए हमें कुल बहुत अधिक परिवारों से जान पड़ताल करनी होगी । यह कोई सतोपजनक तरीका नहीं है ।

वास्तव में जो तरीका अपनाया जाता है वह निम्नलिखित है । परिवारों के एक बड़े प्रतिदर्श को चुना जाता है । इन में से प्रत्येक के लिए कुल आय X और कपड़े पर खर्च Y को मालूम किया जाता है । तब इन प्रेक्षणों के आधार पर X और Y का सबय मालूम किया जाता है ।

§ १५.२ अनुमान के लिए प्रतिरूप (model) का उपयोग

किसी भी Y को X के एक फलन $f(x)$ और एक यादृच्छिक चर ϵ के योग के बराबर मान लिया जाता है।

$$y = f(x) + \epsilon \quad \dots (15.1)$$

यदि $X=x$ दिया हो तो Y का अनुमान $y=f(x)$ लिया जाता है। इस अनुमान के अच्छे होने का निष्पत्ति (criterion) यह है कि $\sum [y-f(x)]^2$ न्यूनतम हो जहाँ यह योग प्रतिदर्श की प्रत्येक इकाई के लिए किया गया हो।

$$\text{समीकरण} \quad E(Y|X=x) = f(x) \quad \dots (15.2)$$

को हम X के ऊपर Y का समाश्रयण कहते हैं। यदि $f(x)$ पर कोई नियंत्रण न रखा जाय तो यह एक बहुत जटिल फलन हो सकता है। यह संभव है कि इस प्रकार के किसी जटिल फलन के लिए प्रतिदर्श में y और $f(x)$ का अंतर शून्य रह जाय, परन्तु यह आवश्यक नहीं कि यह समष्टि के लिए भी सर्वोत्तम होगा। इस दशका के कारण हम प्रायः सरल समाश्रयण के प्रतिरूप (model) से आरम्भ करते हैं। फिर हम उससे कुछ जटिल फलन का आसजन करके देख सकते हैं कि क्या भ्रुटि वर्ग-योग में इस जटिलता के कारण कोई विशेष कमी हुई है। यदि कमी साधारण हो तो हम सरल प्रतिरूप को जटिल प्रतिरूप से उत्तम समझेंगे और उसी के अनुसार अनुमान लगायेंगे।

किस सरल प्रतिरूप से आरम्भ किया जाय यह प्रायः लेखाचित्र (graph) देखकर समझा जा सकता है। बहुधा यह सबब केवल एक-घातीय (linear) ही होता है। यानी

$$y = a + bx + \epsilon \quad \dots (15.3)$$

a और b इस प्रतिरूप के प्राचल हैं। हमारा उद्देश्य a और b को इस प्रकार चुनना है कि $\sum \epsilon = 0$ और $\sum \epsilon^2$ न्यूनतम हो।

§ १५.३ अवकल कलन के कुछ सूत्र

यदि आपने अवकल कलन (differential calculus) का कुछ अध्ययन किया हो तो आपको यह ज्ञात होगा कि यदि $a=a'$ के लिए $g(a,b)$ का मान न्यूनतम है तो

$$\left. \frac{\partial g}{\partial a} \right]_{a=a'} = 0$$

इसी प्रकार यदि $b=b'$ के लिए $g(a,b)$ का मान न्यूनतम हो तो $\left. \frac{\partial g}{\partial b} \right]_{b=b'} = 0$ ।

इन दोनों समीकरणों के हल से हमें a' और b' प्राप्त हो जायेंगे।

यहाँ हम कुछ सूत्र अवकल-कलन के दे रहे हैं जिससे आपको वक्र-आसंजन में सहायता मिलेगी।

(1) यदि $\phi(x) = f_1(x) + f_2(x) + \dots + f_k(x)$
 तो $\frac{\partial \phi(x)}{\partial x} = \frac{\partial f_1(x)}{\partial x} + \frac{\partial f_2(x)}{\partial x} + \dots + \frac{\partial f_k(x)}{\partial x}$ (1')

(2) यदि C एक अचर (constant) है तो
 $\frac{\partial C}{\partial x} = 0$ (2')

(3) यदि $\phi(x) = kx^n$
 तो $\frac{\partial \phi(x)}{\partial x} = knx^{n-1}$ (3')

जहाँ k और n दो अचर हैं।

§ १५.४ एक-घात प्रतिरूप का आसंजन

इन तीन सूत्रों की सहायता से हम एक घात-प्रतिरूप का आसंजन करेंगे।

हमारी समस्या है $\sum_{i=1}^n e_i^2$ को a और b के लिए न्यूनतम करना।

$$\begin{aligned} \sum_{i=1}^n e_i^2 &= \sum_{i=1}^n (y_i - a - bx_i)^2 \\ &= \sum_{i=1}^n y_i^2 + na^2 + b^2 \sum_{i=1}^n x_i^2 \\ &\quad - 2a \sum_{i=1}^n y_i - 2b \sum_{i=1}^n x_i y_i + 2ab \sum_{i=1}^n x_i \end{aligned} \quad \dots\dots (15.4)$$

$$\frac{\partial \sum_{i=1}^n e_i^2}{\partial a} = 2na - 2 \sum_{i=1}^n y_i + 2b \sum_{i=1}^n x_i$$

a के जिस मान के लिए $\sum_{i=1}^n e_i^2$ न्यूनतम होगा उसके लिए

$$\frac{\partial \sum_{i=1}^n e_i^2}{\partial a} = 0 \quad \text{अथवा} \quad \bar{y} = a + b\bar{x} \quad \dots\dots(A)$$

इसी प्रकार $\sum_{i=1}^n e_i^2$ को b द्वारा अवकलित करके हमें निम्नलिखित समीकरण प्राप्त होता है

$$\sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 \quad \dots\dots (B)$$

(A) और (B) को हल करने पर हम देखते हैं कि

$$b = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \quad \dots\dots(15.5)$$

$$= r \frac{\sigma_y}{\sigma_x}$$

b के इस मान को समीकरण (A) में रखने पर

$$a = \bar{y} - \frac{r \sigma_y}{\sigma_x} \bar{x} \quad \dots\dots(15.6)$$

अब यदि हमें X का कोई मान x दिया जाय तो उसके लिए इस रेखा पर Y का मान होगा

$$\left(\bar{y} - \frac{r \sigma_y}{\sigma_x} \bar{x} \right) + \frac{r \sigma_y}{\sigma_x} x = \bar{y} + \frac{r \sigma_y}{\sigma_x} (x - \bar{x})$$

और

यही उस X के लिए Y का अनुमान है।

पिछले अध्याय में जिस सारणी से सह-संबन्ध-गुणांक का परिकलन किया गया था उसके लिए

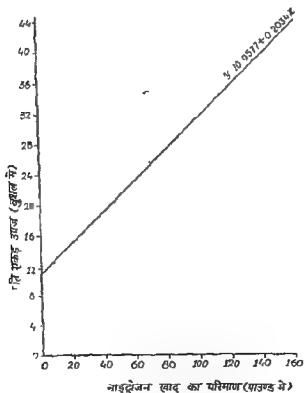
$$b = 0.8422 \times \sqrt{\frac{945.7526}{648.9949}} \times \frac{4}{20}$$

$$\begin{aligned} \text{क्योंकि } \sigma_y^2 &= 945\ 7526 \quad \sigma_x^2 = 648\ 9949 \quad \text{और } \sigma_{xy} = 4\sigma_x \quad \sigma_y = 20\sigma_x \\ &= 0.8422 \times 1\ 2073 \times 0\ 2000 \\ &= 0\ 2034 \end{aligned}$$

$$\begin{aligned} a &= (22 + 4 \times 0\ 7938) - 0\ 2034(70 - 0\ 1003) \\ &= 27\ 1752 - 14\ 2175 = 10\ 9577 \end{aligned}$$

$$\bar{y} = 4\bar{x} + 22, \quad \bar{x} = 20\bar{x}_1 + 70$$

(देखिए, सारणी सत्या 14.2 और § १४.१०)



चित्र ३५—सारणी 14.2 के लिए प्रकीर्ण चित्र और सरल समाधायक रेखा

§ १५.५ अधिक सरल प्रतिरूप

जैसा कि हम पहिले कई बार कह चुके हैं, विज्ञान का एक महत्त्वपूर्ण कार्य है सपूर्ण ज्ञान को कुछ सिद्धांतों अथवा सूत्रों के रूप में रखना। इसके लिए वैज्ञानिक का यह प्रयत्न रहता है कि जहाँ तक हो सके सिद्धांतों को सरल बनाया जाय। यदि वास्तविकता एक सरल सिद्धांत द्वारा समझी जा सकती है तो उसे जटिल बनानेकी कोई आवश्यकता नहीं है।

X के ऊपर Y के समाश्रयण को मालूम करने में भी यह प्रयत्न रहता है कि जितने कम प्राचलों का उपयोग हो उतना ही अच्छा। ऊपर हमने a और b दो प्राचलों का उपयोग किया था। भाप यह जानना चाहेंगे कि क्या नीचे दिये हुए सरल समीकरणों का उपयोग मधैष्ट नहीं था।

$$(i) \quad y = a' + \epsilon \quad \dots (15.7)$$

$$(ii) \quad y = b'x + \epsilon \quad \dots (15.8)$$

आइए, पहिले हम इन समीकरणों के प्राचलों a' और b' का प्राक्कलन करें।

$$(i) \quad \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - a')^2$$

$$= \sum_{i=1}^n y_i^2 - 2a' \sum_{i=1}^n y_i + na'^2 \quad \dots (15.9)$$

$$\frac{\partial \sum_{i=1}^n \epsilon_i^2}{\partial a'} = -2 \sum_{i=1}^n y_i + 2a'n = 0$$

$$\text{अथवा } a' = \bar{y} \quad \dots (15.10)$$

$$(ii) \quad \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - b'x_i)^2$$

$$= \sum_{i=1}^n y_i^2 - 2b' \sum_{i=1}^n x_i y_i + b'^2 \sum_{i=1}^n x_i^2 \quad \dots (15.11)$$

$$\frac{\partial \sum_{i=1}^n \epsilon_i^2}{\partial b'} = -2 \sum_{i=1}^n x_i y_i + 2b' \sum_{i=1}^n x_i^2 = 0$$

अथवा
$$b' = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \quad \dots (15.12)$$

§ १५.६ प्राक्कलकों के प्रसरण

अब हमें यह देना है कि इन सरल प्रतिस्त्रो के लिए त्रुटि के वर्ग-योग क्या है। क्या वे समाश्रयण $y = a + bx + \epsilon$ की त्रुटि के वर्ग-योग से बहुत अधिक है? यदि ऐसा है तो $y = a + bx + \epsilon$ को ही उचित समझा जायगा। यदि ये लगभग बराबर ही हों तो अनेकाहृत सरल प्रतिस्त्रो को चुना जायगा। इसके लिए निम्नलिखित परि-
कलनाओं का परीक्षण किया जाता है

(i) $b = 0$ (ii) $a = 0$

परन्तु इससे पहिले कि हम इन परिकल्पनाओं के परीक्षण का अध्ययन करें, हमें यह जानना आवश्यक है कि यह परीक्षण किन अभिधारणाओं पर आधारित है। ये अभिधारणाएँ निम्नलिखित हैं।

(क) $E(\epsilon | x) = 0$

(ख) $V(\epsilon | x) = \sigma^2_{y,x}$ जो x से स्वतंत्र है

(ग) ϵ का वटन X के किसी भी मान के लिए प्रतापान्य है।

$\sigma^2_{y,x}$ का एक उचित प्राक्कलक $s^2_{y,x}$ है जहाँ

$$\begin{aligned} s^2_{y,x} &= \frac{\sum_{i=1}^n (y_i - a - bx_i)^2}{n-2} \\ &= \frac{\sum_{i=1}^n y_i^2 - a \sum_{i=1}^n y_i - b \sum_{i=1}^n x_i y_i}{n-2} \quad \dots (15.13) \end{aligned}$$

(देखिए, समीकरण (A) और समीकरण (B))

ऊपर जिस सारणी के लिए हमने सह-संबन्ध-गुणांक का परिकल्पन किया था उसके लिए X पर Y के समाश्रयण रेखा का समीकरण था

$$y = a + bx = 10\ 9577 + 0\ 2034x$$

क्योंकि हम ऊपर देल चुके हैं कि $a = 10\ 9577$
तथा $b = 0\ 2034$ और $y_i = (4y'_i + 22)$, $x_i = 20x'_i + 70$

$$\sum_{i=1}^{194} y_i = (154 \times 4) + (22 \times 194) = 4884$$

$$\begin{aligned} \sum_{i=1}^{194} y_i^2 &= (1068 \times 16) + (2 \times 22 \times 4 \times 154) + (22 \times 22 \times 194) \\ &= 138,088 \end{aligned}$$

$$\begin{aligned} \sum_{i=1}^{194} x_i y_i &= (659 \times 80) + (280 \times 154) + (440 \times -1) \\ &\quad + 22 \times 70 \times 194 \\ &= 394,160 \end{aligned}$$

(देखिए, सारणी सख्या 142 और § १४ १०)

इसलिए इन आंकड़ों के लिए

$$\begin{aligned} s_{y,x}^2 &= \frac{138\ 088 - (10\ 9577 \times 4884) - (0\ 2034 \times 394\ 160)}{194 - 2} \\ &= \frac{4398\ 4492}{192} \\ &= 22\ 9086 \end{aligned}$$

यदि n प्रेक्षण-युग्मों के अनेक प्रतिदर्श एक ऐसी समष्टि में से चुने जायें जिसका सरल समाश्रयण प्रतिरूप उचित ही और यदि स्वतंत्र चर X के मान x_1, x_2, \dots, x_n सब प्रतिदर्शों के लिए समान हों तो

(1) b के प्राक्कलक \hat{b} का माध्य b होगा
यानी $E(\hat{b}) = b$

(15 14)

(2) \hat{b} का प्रसरण निम्नलिखित होगा

$$V(\hat{b}) = \frac{\sigma_y^2 x}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (15.15)$$

$$(3) E(a) = a \quad (15.16)$$

$$(4) V(a) = \frac{\sigma_y^2 y \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \quad (15.17)$$

§ १५.७ परिकल्पना परीक्षण

यदि प्रतिदश-परिमाण बहुत बड़ा हो तो ऊपर लिखे हुए अनुबन्धा के अनुसार \hat{b} के प्रतिदर्शज वटन (sampling distribution) का ऐसे प्रसामान्य वटन द्वारा सन्निकटन किया जा सकता है जिसका माध्य b और प्रसरण $\frac{\sigma_y^2 x}{\sum_{i=1}^n (x_i - \bar{x})^2}$ हो।

σ_y^2 अज्ञात है परन्तु इस बड़े प्रतिदर्श में σ_y^2 के स्थान पर उसके प्राक्कलक s_y^2

का उपयोग किया जा सकता है। इसलिए यदि \hat{b} का मान $-1.96 \sqrt{\frac{s_y^2 x}{\sum_{i=1}^n (x_i - \bar{x})^2}}$

से कम अथवा $+1.96 \sqrt{\frac{s_y^2 x}{\sum_{i=1}^n (x_i - \bar{x})^2}}$ से अधिक हो तो हम निराकरणिय परि-

कल्पना $b=0$ को पाच प्रतिशत स्तर पर अस्वीकार कर सकते हैं। इसी प्रकार \hat{a} के वटन का सन्निकटन एक प्रसामान्य वटन से किया जा सकता है जिसके माध्य और प्रसरण समीकरण (15.16) तथा (15.17) से प्राप्त होते हैं। इसलिए यदि \hat{a}

का परिकल्पित मान $-1.96 \sqrt{\frac{s_y^2 y \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}}$ से कम हो अथवा

$+ 1.96 \sqrt{\frac{s_y^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}}$ से अधिक हो तो हम नियन्त्रणीय परिवर्तन

$n=0$ को पाँच प्रतिशत-स्तर पर अस्वीकार कर देते हैं। प्रेक्षित मान-युग्मों द्वारा हमें इस बात का आभास मिल सकता है कि संघट्टि में सरल समाश्रयण का प्रतिरूपण कहाँ तक उपयुक्त है परन्तु यह आभास हमें प्रेक्षित मानों के परास के लिए ही मिल सकता है। यह बहुत संभव है कि प्रेक्षित परास में तो सरल समाश्रयण उपयुक्त हो, परन्तु परास के बाहर समाश्रयण का रूप कुछ और ही हो। इस कारण प्रेक्षण के आधार पर प्रेक्षित परास के बाहर के किसी मान के लिए मानों के माध्य का अनुमान जरा सोच समझ कर ही लगाना चाहिए।

§ १५.८ द्वि-घाती परवलय का आसजन

यदि समाश्रयण वक्र का समीकरण एक घात फलन हो तो हम देख चुके हैं कि प्रतिदर्शों से हम समाश्रयण वक्र के प्राचलों का प्राक्कलन किस प्रकार करते हैं। यही विधि बहुघाती परवलय-वक्रों में समाश्रयण होने पर भी अपनायी जाती है। द्वि-घाती परवलय (parabola of second degree) के प्राचलों के प्राक्कलन की विधि उदाहरण स्वरूप नीचे दी हुई है।

द्वि-घाती परवलय का समीकरण निम्नलिखित होता है।

$$y = a + bx + cx^2 \quad \dots (15.18)$$

a , b और c इस वक्र के प्राचल हैं। यदि प्रतिदर्शों में (X, Y) युग्म के मान (x_1, y_1) , (x_2, y_2) , ..., (x_n, y_n) हो तो हम a , b और c के ऐसे मान माफूम करना चाहते हैं जिनके लिए

$$Q = \sum_{i=1}^n (y_i - a - bx_i - cx_i^2)^2$$

न्यूनतम हो।

$$Q = \sum_{i=1}^n y_i^2 + na^2 + b^2 \sum_{i=1}^n x_i^2 + c^2 \sum_{i=1}^n x_i^4$$

$$\begin{aligned}
 & -2a \sum_{i=1}^n y_i - 2b \sum_{i=1}^n x_i y_i - 2c \sum_{i=1}^n x_i^2 y_i \\
 & + 2ab \sum_{i=1}^n x_i + 2ac \sum_{i=1}^n x_i^2 + 2bc \sum_{i=1}^n x_i^3 \dots (15.19)
 \end{aligned}$$

a के जिस मान के लिए न्यूनतम होगा वह निम्नलिखित समीकरण को सतुष्ट करेगा।

$$\frac{\partial Q}{\partial a} = 0$$

$$\text{अथवा } 2an - 2 \sum_{i=1}^n y_i + 2b \sum_{i=1}^n x_i + 2c \sum_{i=1}^n x_i^2 = 0$$

$$\text{अथवा } \sum_{i=1}^n y_i = na + b \sum_{i=1}^n x_i + c \sum_{i=1}^n x_i^2 \dots (A)$$

इसी प्रकार b और c के लिए हमें निम्नलिखित समीकरण प्राप्त होंगे

$$\sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 + c \sum_{i=1}^n x_i^3 \dots (B)$$

$$\sum_{i=1}^n x_i^2 y_i = a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i^3 + c \sum_{i=1}^n x_i^4 \dots (C)$$

a , b और c प्राचलों में (A), (B) और (C) तीन युगपद (simultaneous) समीकरण हैं। इनके हल से हमें a , b , और c के इच्छित मान ज्ञात हो जाते हैं।

(A) और (B) में से a का निरसन (elimination) करने से हमें निम्नलिखित समीकरण प्राप्त होता है।

$$\left[\sum_{i=1}^n x_i y_i - \bar{x} \sum_{i=1}^n y_i \right] = b \left[\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i \right] + c \left[\sum_{i=1}^n x_i^3 - \bar{x} \sum_{i=1}^n x_i^2 \right]$$

$$\text{अथवा } S_{xy} = b S_{xx} + c S_{x^2x} \dots (D)$$

$$\text{जहाँ } S_{x_1x_2} = \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)$$

इसी प्रकार (A) और (C) में से a का निरसन करने से हमें निम्नलिखित समीकरण प्राप्त होता है।

$$S_{xy} = b S_{xx} + c S_{xz} \quad \dots (E)$$

(D) को S_{xy} तथा (E) को S_{xz} से गुणा करके एक में से दूसरे घटाने पर हमें निम्नलिखित समीकरण मिलता है

$$S_{xy} S_{xz} - S_{xz} S_{xy} = C \{ [S_{xz}]^2 - S_{xx} S_{zz} \}$$

$$\therefore C = \frac{S_{xy} S_{xz} - S_{xz} S_{xy}}{[S_{xz}]^2 - S_{xx} S_{zz}} \quad \dots (C')$$

c के इस मान को (D) में निविष्ट करने पर

$$b = \frac{S_{xy} S_{zz} - S_{xz} S_{yz}}{[S_{xz}]^2 - S_{xx} S_{zz}} \quad \dots (B')$$

b और c के इन मानों को समीकरण (A) में रखकर हम a का मान प्राप्त कर सकते हैं।

$$a = \bar{y} - b\bar{x} - c\bar{z} \quad \dots (A')$$

यदि आपकी इच्छा हो तो जिस सारणी का उपयोग अभी तक हम करते आ रहे हैं उसके लिए a , b और c का परिकलन ऊपर दी हुई विधि से कर सकते हैं।

अध्याय १६

प्रतिबंधी घटन, सह-संबंधानुपात और माध्य वर्ग आसंग

(Conditional Distribution, Correlation Ratio and Mean Square Contingency)

प्रतिबंधी प्रायिकता (conditional probability) से आप परिचित ही है। आप जानते हैं कि यदि A और B दो घटनाएँ हो तो यह दिये होने पर कि B घटी है A की प्रायिकता निम्नलिखित सूत्र से प्राप्त होती है

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

§ १६.१ असतत चर

अब मान लीजिए कि (X, Y) एक असतत द्वि-चर है तथा X और Y क्रमशः x_1, x_2, \dots, x_m तथा y_1, y_2, \dots, y_n मान धारण करते हैं। बिंदु (x_i, y_j) पर जो प्रायिकता है उसे हम P_{ij} से सूचित करेंगे।

$$P[X = x_i, Y = y_j] = p_{ij} \quad \dots\dots(16.1)$$

यदि हम p_i द्वारा $X=x_i$ होने की प्रायिकता को सूचित करें तो

$$p_i = P[X=x_i] = \sum_{k=1}^n p_{ik} \quad \dots\dots(16.2)$$

$$\therefore P(Y=y_j | X=x_i) = \frac{P(X=x_i, Y=y_j)}{P(X=x_i)} = \frac{p_{ij}}{p_i} \quad (16.3)$$

यदि हम $X=x_i$ के दिये होने पर Y के प्रत्येक मान के लिए प्रतिबंधी प्रायिकता मालूम करें तो $X=x_i$ के दिये होने पर Y का प्रतिबंधी घटन (conditional distribution) प्राप्त होता है। यह स्पष्ट है कि यह प्रतिबंधी घटन केवल उसी दशा में अर्थ-पूर्ण हो सकता है जब p_i न्यून न हो।

प्रतिबंधी माध्य—प्रतिबंध $X=x_i$ के दिये होने पर (X, Y) के किसी फलन $\phi(x, y)$ का माध्य निम्नलिखित रूप से प्राप्त किया जा सकता है।

$$E[\phi(X, Y) | X=x_i] = \sum_{k=1}^n \phi(x_i, y_k) \frac{p_{ik}}{p_i}$$

$$= \frac{\sum_{k=1}^n \phi(x_i, y_k) p_{ik}}{p_i} \quad \dots (164)$$

यदि $\phi(X, Y) = Y$ तो

$$E(Y | X=x_i) = \frac{\sum_{k=1}^n y_k p_{ik}}{p_i} \quad \dots (165)$$

इस माध्य को Y का प्रतिबंधी माध्य कहते हैं और इसको $m_Y^{(i)}$ से सूचित करते हैं। यदि $\phi(X, Y) = [Y - m_Y^{(i)}]^2$ हो तो हॉ Y का प्रतिबंधी प्रसरण प्राप्त होता है।

$$V(Y | X=x_i) = \frac{\sum_{k=1}^n (y_k - m_Y^{(i)})^2 p_{ik}}{p_i} \quad \dots (166)$$

इसी प्रकार प्रतिबंध $Y=y_k$ से संबंधित X का बंटन, उसका माध्य और प्रसरण भी हम मालूम कर सकते हैं।

§ १६२ सतत चर

यदि (X, Y) का बंटन सतत हो और $f(x, y)$ उसका घनत्व फलन हो तो

$$P[x < X < x+h] = \int_{-\infty}^{x+h} \int_{-\infty}^{\infty} f(x, y) dx dy$$

यदि प्रतिबंध $(x < X < x+h)$ दिया हो तो $Y < y$ की प्रतिबंधी प्रायिकता निम्नलिखित होगी

$$\begin{aligned}
 P [Y \leq y | x < X < x+h] &= \frac{P [Y \leq y \text{ व } x < X < x+h]}{P [x < X < x+h]} \\
 &= \frac{\int_x^{x+h} \int_{-\infty}^y f(x, y) dx dy}{\int_x^{x+h} \int_{-\infty}^{\infty} f(x, y) dx dy} \quad (167)
 \end{aligned}$$

यदि $X=x$ पर X के घटन का घनत्व फलन $f_1(x)$ धनात्मक है तो

$$\begin{aligned}
 \lim_{h \rightarrow 0} P (Y \leq y | x < X < x+h) &= \frac{\int_{-\infty}^y f(x, y) dy}{\int_{-\infty}^{\infty} f(x, y) dy} \\
 &= \frac{\int_{-\infty}^y f(x, y) dy}{f_1(x)} \quad (168)
 \end{aligned}$$

प्रतिबंध $X=x$ के लिए यह Y का प्रतिबन्धी घटन फलन (conditional distribution function) कहलाता है। इस फलन का y के प्रति अवकलन (differentiate) करने पर हम Y का प्रतिबन्धी घनत्व फलन $f_2(y|x)$ प्राप्त होता है।

$$f_2(y|x) = \frac{f(x, y)}{f_1(x)} \quad (169)$$

प्रतिबन्धी माध्य— $X=x$ दिए होने पर (XY) के किसी फलन $\phi(XY)$ का प्रतिबन्धी माध्य निम्नलिखित होगा।

$$\begin{aligned}
 E[\phi(XY)|X=x] &= \frac{\int_{-\infty}^{\infty} \phi(x, y) f_2(y|x) dy}{\int_{-\infty}^{\infty} \phi(x, y) f(x, y) dy} \\
 &= \frac{\int_{-\infty}^{\infty} \phi(x, y) f(x, y) dy}{f_1(x)} \quad (1610)
 \end{aligned}$$

Y के प्रतिबन्धी माध्य को यदि हम $m_2(x)$ से और प्रतिबन्धी प्रसरण को $\sigma_2^2(x)$ से सूचित करें तो

$$m_2(x) = E(Y|X=x) = \frac{\int_{-\infty}^{\infty} y f(x,y) dy}{f_1(x)} \dots (16.11)$$

$$\sigma_2^2(x) = V(Y|X=x) = \frac{\int_{-\infty}^{\infty} [y - m_2(x)]^2 f(x,y) dy}{f_1(x)} \dots (16.12)$$

इसी प्रकार X के प्रतिबन्धी बटन, प्रतिबन्धी माध्य $m_1(y)$ और प्रतिबन्धी प्रसरण $\sigma_1^2(y)$ की व्याख्या की जा सकती है।

§ १६३ समाश्रयण (Regression)

$m_2(x)$ स्पष्टतः x का एक फलन है। x के विभिन्न मानों के लिए यह विभिन्न मान धारण कर सकता है। $y = m_2(x)$ एक वक्र का समीकरण है जो (X, Y) समतल में x के विभिन्न मानों के लिए $[x, m_2(x)]$ बिन्दुओं को मिलाता है। इस वक्र को X पर Y का समाश्रयण कहते हैं। इसी प्रकार Y के विभिन्न मानों के लिए $[m_1(y), y]$ बिन्दुओं को मिलाता हुआ वक्र $x = m_1(y)$ है जो Y पर X का समाश्रयण कहलाता है। यदि $m_2(x)$ x का एक-घात फलन (linear function) होता है तो X पर Y के समाश्रयण को सरल समाश्रयण कहते हैं। इसके प्राचलों का प्राक्कलन प्रतिदर्श के आधार पर कैसे किया जाता है, यह हम पिछले अध्याय में लिख ही चुके हैं।

समाश्रयण वक्रों का एक महत्वपूर्ण गुण होता है। X के सब फलनों में से यदि हम उस फलन $\phi(x)$ को चुनें जिसके लिए $E[Y - \phi(x)]^2$ न्यूनतम हो तो यह सिद्ध किया जा सकता है कि $\phi(x) = E(y|x)$ क्योंकि

$$\begin{aligned} E[Y - \phi(x)]^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [y - \phi(x)]^2 f(x,y) dx dy \\ &= \int_{-\infty}^{\infty} f_1(x) dx \int_{-\infty}^{\infty} [y - \phi(x)]^2 f_2(y|x) dy \dots (16.13) \end{aligned}$$

आप यह जानते ही हैं कि किसी भी वंटन के लिए $(y-a)^2$ का प्रत्याशित मान $a=E(y)$ पर न्यूनतम होता है। इसलिए Y के प्रतिबंधी वंटन के लिए $[y-\phi(x)]^2$ का प्रत्याशित मान $\phi=E(Y|x)$ होने पर न्यूनतम होगा। इस प्रकार X पर Y का समाश्रयण वक्र ऐसा होता है कि X के ज्ञान के आधार पर Y का अनुमान लगाने के लिए यदि इस वक्र पर ज्ञात x के लिए y स्थानांक (coordinate) फो लें तो वृद्धि $[y-\phi(x)]$ के वर्ग का प्रत्याशित मान अन्य किसी भी वक्र पर आधारित अनुमान की वृद्धि के वर्ग के प्रत्याशित मान से कम होगा।

§ १६.४ सह-संबंधानुपात (Correlation ratio)

यदि Y के माध्य को m_2 और प्रसरण को σ_2^2 से सूचित किया जाय तो

$$\begin{aligned}\sigma_2^2 &= E(Y-m_2)^2 \\ &= E[Y-m_2(X)+m_2(X)-m_2]^2 \\ &= E[Y-m_2(x)]^2 + E[m_2(X)-m_2]^2\end{aligned}\quad (16.14)$$

इस प्रकार हम देखते हैं कि Y के प्रसरण को दो सघटकों (components) के रूप में रखा जा सकता है। एक सघटक तो उसके प्रतिबंधी माध्य $m_2(X)$ से Y का माध्य वर्ग विचलन है और दूसरा $m_2(x)$ का उसके माध्य m_2 से माध्य वर्ग विचलन।

यदि हम $\frac{E[m_2(X)-m_2]^2}{\sigma_2^2}$ को η^2 द्वारा सूचित करें तो

$$\begin{aligned}\eta^2 &= \frac{E[m_2(X)-m_2]^2}{\sigma_2^2} \\ &= 1 - \frac{E[Y-m_2(x)]^2}{\sigma_2^2}\end{aligned}\quad (16.15)$$

$$\therefore 1 - \eta^2 = \frac{E[Y-m_2(x)]^2}{\sigma_2^2} \geq 0$$

$$\therefore 0 \leq \eta^2 \leq 1 \quad \dots (16.16)$$

इस मान η को हम सह-संबंधानुपात कहते हैं। यदि समाश्रयण एक-घाती है तो

$$m_2(x) = a + bx \text{ और}$$

$$1 - \eta^2 = \frac{E [Y - a - bx]^2}{\sigma_2^2}$$

$$= 1 - \rho^2$$

इसलिए इस दशा में $\eta^2 = \rho^2$

यह स्पष्ट है कि $\eta^2 = 1$ केवल उसी अवस्था में हो सकता है जब कि $E [Y - m_2(x)] = 0$ हो, अर्थात् जब Y के $m_2(x)$ में भिन्न होने की प्रायिकता शून्य हो। η^2 को इस कारण प्रायिकताओं की समाख्ययण चक्र के पास एकत्रित होने की प्रवृत्ति का एक माप समझा जा सकता है।

जिस प्रकार सतत चर के लिए सह-संबंधानुपात की व्याख्या की गयी है उसी प्रकार असतत चर-युग्म के लिए भी की जा सकती है।

इस दशा में

$$\eta^2 = \frac{1}{\sigma_2^2} E [m_2^{(j)} - m_2]^2$$

$$= \frac{1}{\sigma_2^2} \sum_{j=1}^m (m_2^{(j)} - m_2)^2 p_j \quad \dots\dots(16.17)$$

§ १६५ माध्य वर्ग आसंग

सह-संबंधानुपात हमें X पर Y की निर्भरता का आभास देता है। इसी उद्देश्य से अनेक अन्य मापों का भी प्रस्ताव किया गया है जिनमें से एक माध्य वर्ग आसंग (mean square contingency) है। इसका उपयोग केवल असतत समष्टियों के लिए किया जाता है।

यदि असतत चर युग्म का घटन निम्नलिखित है

$$P [X = x_j, Y = y_k] = p_{jk}; \quad j = 1, 2, \dots, m; \quad k = 1, 2, \dots, n$$

तो हम इन प्रायिकताओं को एक मारणी में रख सकते हैं जिसमें m पंक्तियाँ और n स्तम हैं।

सारणी सख्या 16 1

[X, Y] का वटन

X \ Y		Y ₁	Y ₂		Y _k		Y _n	योग
		(1)	(2)		(k)		(n)	
x ₁	(1)	P ₁₁	P ₁₂		P _{1k}		P _{1n}	P ₁
x ₂	(2)	P ₂₁	P ₂₂		P _{2k}		P _{2n}	P ₂
x _i	(i)	P _{i1}	P _{i2}		P _{ik}		P _{in}	P _i
x _m	(m)	P _{m1}	P _{m2}		P _{mk}		P _{mn}	P _m
योग		P ₁	P ₂		P _k		P _n	I

क्योंकि हम इस सारणी में से हम प्रकार की पक्तियों या स्तंभों को छोड़ सकते हैं जिनमें सब प्रायिकताएँ शून्य हों, इसलिए प्रत्येक पक्ति का योग p_i और स्तंभ का योग p_k शून्य से अधिक होगा। इस दशा में वटन के माध्य-वर्ग आसग की—जिसको ϕ^2 से सूचित किया जाता है—निम्नलिखित परिभाषा है

$$\begin{aligned} \phi^2 &= \sum_{i=1}^m \sum_{k=1}^n \frac{(P_{ik} - P_i P_k)^2}{P_i P_k} \\ &= \sum_{i=1}^m \sum_{k=1}^n \frac{P_{ik}^2}{P_i P_k} - 1 \end{aligned} \quad (16 18)$$

ϕ^2 शून्य केवल उस स्थिति में हो सकता है जब प्रत्येक युग्म (i, k) के लिए $p_{ik} = p_i p_k$ परंतु हम जानते हैं कि इस दशा में दोनों चर स्वतंत्र होते हैं। इसके अतिरिक्त $p_{ik} \leq p_i$ और $p_{ik} \leq p_k$ होने के कारण

$$\sum_{i=1}^m \sum_{k=1}^n \frac{p_{ik}^2}{p_i p_k} \leq \sum_{i=1}^m \sum_{k=1}^n \frac{p_{ik}}{p_k} = n \quad (16.19)$$

$$\text{और } \sum_{i=1}^m \sum_{k=1}^n \frac{p_{ik}}{p_i p_k} \leq \sum_{i=1}^m \sum_{k=1}^n \frac{p_{ik}}{p_i} = m \quad (16.20)$$

$$\therefore \phi^2 \leq q-1$$

$$\text{जहाँ } q = M_{1n}(m, n)$$

$M_{1n}(m, n)$ से हमारा तात्पर्य m और n संख्याओं में से छोटी वाली संख्या से है।

इस प्रकार $0 \leq \frac{\phi^2}{q-1} \leq 1$ और $\frac{\phi^2}{q-1}$ का उपयोग दोनों चरों की

पारस्परिक निर्भरता के एक मानकित मापनी (standardized scale) पर लिये हुए माप के लिए किया जा सकता है।

भाग ४

प्राक्कलन

प्राक्कलन के आरंभिक सिद्धान्त

(Elementary Principles of Estimation)

§ १७.१ प्राक्कलक और उसके कुछ इच्छित गुण

समाधयण के अध्यायों में हम कुछ समष्टि प्राचलों का प्राक्कलन कर चुके हैं। इसी प्रकार परिकल्पना परीक्षण में—विशेष रूप से x^2 -परीक्षण में—हम प्राचलों के प्राक्कलन से कुछ परिचय प्राप्त कर चुके हैं। किसी भी प्राचल का प्राक्कलन करने के लिए प्रेक्षणों के एक फलन की आवश्यकता होती है जिसे प्राक्कलक (estimator) कहते हैं।

इस अध्याय में हम यह देखेंगे कि प्राक्कलकों को प्राप्त करने की साधारण विधियाँ क्या हैं और किस प्रकार के प्राक्कलकों की अच्छा समझा जाता है।

किसी प्राचल का प्राक्कलक क्या होना चाहिए, यह पूर्णतः स्पष्ट नहीं है। यद्यपि समष्टि के माध्य के लिए प्रतिदर्श-माध्य को प्राक्कलक मानना स्पष्टतया उचित जान पड़ता है, परंतु समष्टि-प्रसरण का प्राक्कलक प्रतिदर्श-प्रसरण नहीं होता। उसमें हमें प्रतिदर्श के माध्य से प्राचल के विचलनों के वर्ग-योग को प्रतिदर्श परिमाण से एक कम संख्या द्वारा भाग देना होता है। ऐसा क्यों किया जाता है इसका कारण आप अवश्य जानना चाहेंगे। आप यह भी जानना चाहेंगे कि किराी नवीन स्थिति में जिससे आप अभी तक परिचित नहीं हैं, प्राचल का प्राक्कलन किस प्रकार किया जायगा।

यदि हम समष्टि से एक यादृच्छिक प्रतिदर्श x_1, x_2, \dots, x_n चुनें तो इन मानों के किसी भी फलन $g(x_1, x_2, \dots, x_n)$ को समष्टि के किसी प्राचल θ का प्राक्कलक माना जा सकता है। एक उत्तम प्राक्कलक के लिए हम चाहेंगे कि

$|g(x_1, x_2, \dots, x_n) - \theta|$ जहाँ तक हो सके छोटा हो। परंतु क्योंकि x_1, x_2, \dots, x_n यादृच्छिक चर हैं इसलिए $|g(x_1, x_2, \dots, x_n) - \theta|$ भी एक यादृच्छिक चर है—अचर नहीं। इस कारण इसके छोटे होने की परिभाषा हमें इसके प्रत्याशित मान (expected value) अथवा इसकी प्रायिकता के रूप में

करनी होगी। इस रूप में प्राक्कलनों के कुछ इच्छित गुणों की परिभाषा हम नीचे दे रहे हैं।

(1) अनभिनतता (Unbiasedness) मान लीजिए कि $g(x_1, x_2, \dots, x_n)$ को हम t_n से सूचित करते हैं। यदि $E[t_n - \theta] = 0$ तो हम t_n को एक अनभिनत प्राक्कलक (unbiased estimator) कहते हैं। किसी प्राक्कलक के अनभिनत होने के गुण को अनभिनतता कहते हैं।

यदि $E[t_n - \theta]$ शून्य के बराबर न हो तो प्राक्कलक अभिनत कहलाता है और तब $E[t_n - \theta]$ को हम $B(t_n)$ से सूचित करते हैं और इसे प्राक्कलक की अभिनति (bias) कहते हैं।

उदाहरण के लिए एक प्रसामान्य वटन $N(\mu, \sigma)$ में से चुने हुए n परिमाण के प्रतिदर्श का माध्य \bar{x}_n वटन के माध्य का एक अनभिनत प्राक्कलक है। क्योंकि \bar{x}_n

एक $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ वर है। $\therefore E(\bar{x}_n) = \mu$, परंतु प्रतिदर्शों का प्रसरण

$s_n^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2$ वटन के प्रसरण σ^2 के लिए अनभिनत नहीं है क्योंकि

$$E(s_n^2) = E\left[\frac{1}{n} \sum_{i=1}^n [(x_i - \mu) - (\bar{x} - \mu)]^2\right] = \frac{1}{n} \left[\sum_{i=1}^n \sigma^2 - \sigma^2 \right] = \frac{n-1}{n} \sigma^2$$

s_n^2 की अभिनति $\frac{n-1}{n} \sigma^2 - \sigma^2 = -\frac{1}{n} \sigma^2$ है।

(2) दक्षता (efficiency)—यदि हम केवल अनभिनत प्राक्कलकों पर विचार करें तो इनमें से एक ऐसा हो सकता है जिसका प्रसरण अन्य सब प्राक्कलकों के प्रसरण से कम हो। इस प्रकार के प्राक्कलक को दक्ष प्राक्कलक (efficient estimator) अथवा न्यूनतम प्रसरण-अनभिनत प्राक्कलक (minimum variance unbiased estimator) कहते हैं। यदि किसी प्राक्कलक t का प्रसरण σ^2 हो और एक दक्ष प्राक्कलक का प्रसरण σ^2 हो तो t की दक्षता (efficiency) को $\frac{\sigma^2}{\sigma^2}$ द्वारा

नापा जाता है। इस दक्षता को $e(t)$ से सूचित करते हैं।

$$e(t) = \frac{\sigma^2}{\sigma^2} \quad (171)$$

यदि t और t' दो अनभिन्नत प्राक्कलक हों तो t को t' से अधिक दक्ष माना जायगा यदि t की दक्षता t' की दक्षता से अधिक हो अथवा $V(t) < V(t')$

मान लीजिए x_1, x_2, \dots, x_n को इस प्रकार क्रम y_1, y_2, \dots, y_n में रखा जाय कि $y_1 < y_2 < \dots < y_n$ । यदि n एक विषम संख्या हो तो $\frac{y_{\frac{n+1}{2}}}{2}$

इन प्रक्षणों की माध्यिका होगी । क्योंकि एक प्रसामान्य $N(\mu, \sigma)$ वटन में माध्य और माध्यिका दोनों μ होते हैं इसलिए यह सिद्ध किया जा सकता है कि इस प्रकार के वटन से चुने हुए यादृच्छिक प्रतिदर्श के लिए

$$E\left(\frac{y_{\frac{n+1}{2}}}{2}\right) = \mu$$

यानी $\frac{y_{\frac{n+1}{2}}}{2}$ भी μ का एक अनभिन्नत प्राक्कलक है । परंतु $V\left(\frac{y_{\frac{n+1}{2}}}{2}\right) > \frac{\sigma^2}{n} =$

$V(\bar{x})$ इसलिए μ के प्राक्कलन के लिए $\frac{y_{\frac{n+1}{2}}}{2}$ से \bar{x}_n अधिक दक्ष है ।

संगति (Consistency)

$P[|t_n - \theta| < \epsilon]$ प्रतिदर्श परिमाण n का एक फलन है । यहाँ ϵ कोई भी निश्चित घनात्मक संख्या है । अधिकतर यह आशा की जाती है कि यह प्रायिकता n के साथ साथ बढ़ती जायगी । यदि किसी प्राक्कलक t_n के लिए n के ∞ की ओर प्रवृत्त होने के साथ यह प्रायिकता 1 की ओर प्रवृत्त हो तो t_n को एक संगत (Consistent) प्राक्कलक कहेंगे । इस प्रकार यदि t_n एक संगत प्राक्कलक है तो $\lim_{n \rightarrow \infty} P[|t_n - \theta| < \epsilon] = 1$ (17.2)

उदाहरण के लिए एक प्रसामान्य वटन $N(\mu, \sigma)$ से चुने हुए प्रतिदर्श का माध्य \bar{x}_n संगत है

$$\begin{aligned} P[|\bar{x}_n - \mu| < \epsilon] &= P\left[-\frac{\epsilon}{\sigma/\sqrt{n}} < \frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}} < +\frac{\epsilon}{\sigma/\sqrt{n}}\right] \\ &= P\left[-\frac{\epsilon}{\sigma} \sqrt{n} < N(0,1) < \frac{\epsilon}{\sigma} \sqrt{n}\right] \end{aligned}$$

$$\therefore \lim_{n \rightarrow \infty} P[|\bar{x}_n - \mu| < \epsilon] = P[-\infty < N(0,1) < +\infty] = 1$$

पर्याप्त (sufficiency) यदि (x_1, x_2, \dots, x_n) के संयुक्त वटन $f(x_1, x_2, \dots, x_n; \theta)$ को निम्नलिखित रूप में रखा जा सके

$$f(x_1, x_2, \dots, x_n; \theta) = f_1(t; \theta) \times f_2(x_1, x_2, \dots, x_n)$$

जहाँ $f_2(x_1, x_2, \dots, x_n)$ ऐसा फलन हो जो θ से स्वतंत्र हो और θ के लिए t एक प्राक्कलक हो तो t को एक पर्याप्त प्राक्कलक (sufficient estimator) कहते हैं और किसी प्राक्कलक के पर्याप्त होने के गुण को पर्याप्त कहते हैं।

यह सिद्ध किया जा सकता है कि यदि t_1 पर्याप्त हो और θ का कोई अन्य प्राक्कलक t_2 हो जो t_1 का फलन नहीं है तो t_1 और t_2 के संयुक्त वटन को निम्नलिखित रूप में रखा जा सकता है

$$\psi(t_1, t_2, \theta) = \psi_1(t_1; \theta) \psi_2(t_2, t_1) \dots \dots (17.3)$$

जहाँ ψ_2 में θ का कोई स्थान नहीं है। इस समीकरण से यह पता चलता है कि t_1 के ज्ञात होने पर t_2 का प्रायिकता घनत्व $\psi_2(t_2, t_1)$ है जो θ से स्वतंत्र है। अर्थात् t_1 के ज्ञात होने पर अन्य कोई भी प्राक्कलक t_2 पर कोई अतिरिक्त प्रकाश नहीं डाल सकता। प्रेक्षण x_1, x_2, \dots, x_n जो कुछ भी सूचना हमें प्राक्कलक के बारे में देते हैं, वह सब हमें प्राक्कलक t_1 से मिल जाती है। यही कारण इसको पर्याप्त कहने का है।

यदि x_1, x_2, \dots, x_n एक $N(\mu, 1)$ में चुने हुए n प्रेक्षण हैं तो $\bar{x} = (x_1, x_2, \dots, x_n)$ का संयुक्त वटन निम्नलिखित है

$$f(\bar{x}, \mu) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2}$$

$$\text{परंतु } \sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)^2$$

$$= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2$$

$$\therefore f(\bar{x}, \mu) = \sqrt{\frac{n}{2\pi}} e^{-\frac{1}{2} \left[\frac{\bar{x} - \mu}{\sqrt{n}} \right]^2} \times \frac{1}{\sqrt{n(2\pi)^{\frac{n-1}{2}}}} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2}$$

इस प्रकार इस सगुणत वटन को दो गुणन खंडों (factors) के गुणन के रूप में रखा जा सकता है जिसमें से पहिला गुणन खंड तो \bar{x} का घनत्व-फलन है और दूसरा गुणन खंड μ से स्वतंत्र है। इसलिए μ के लिए \bar{x} एक घटावित प्राक्कलक है।

§ १७.२ दो अनभिन्नत प्राक्कलकों का संघनन

यदि t_1 और t_2 दोनों एक ही प्राचल θ के अनभिन्नत प्राक्कलक हैं और l_1 तथा l_2 दो ऐसी संख्याएँ हैं जिनका योग 1 है तो $l_1 t_1 + l_2 t_2$ भी 0 का एक अनभिन्नत प्राक्कलक है क्योंकि

$$\begin{aligned} E(l_1 t_1 + l_2 t_2) &= E(l_1 t_1) + E(l_2 t_2) \quad \dots\dots(17.4) \\ &= (l_1 + l_2)0 \\ &= 0 \end{aligned}$$

यदि t_1 का प्रसरण σ_1^2 , t_2 का प्रसरण σ_2^2 तथा t_1 और t_2 का सहसंबंध गुणांक ρ हो तो $V(l_1 t_1 + l_2 t_2) = E[l_1(t_1 - \theta) + l_2(t_2 - \theta)]^2$
 $= l_1^2 \sigma_1^2 + 2l_1 l_2 \rho \sigma_1 \sigma_2 + l_2^2 \sigma_2^2 \quad \dots\dots(17.5)$

इस प्रकार के दो अनभिन्नत प्राक्कलकों का हम इस प्रकार संघनन करना चाहते हैं कि $V(l_1 t_1 + l_2 t_2)$ न्यूनतम हो। इसके लिए निम्नलिखित विधि काम में लायी जाती है।

हम पहिले ही एक नवीन राशि Q की परिभाषा निम्नलिखित समीकरण से करते हैं

$$Q = V(l_1 t_1 + l_2 t_2) - \lambda [l_1 + l_2 - 1] \quad \dots\dots(A)$$

अब हम l_1 और l_2 के वे मान मालूम करते हैं जो Q को न्यूनतम कर देते हैं। इसके लिए हमें निम्नलिखित समीकरण प्राप्त होते हैं—

$$(1) \quad \frac{\partial Q}{\partial l_1} = 0$$

अथवा $2l_1 \sigma_1^2 + 2l_2 \sigma_1 \sigma_2 \rho = \lambda \quad \dots\dots(B)$

तथा $(2) \quad \frac{\partial Q}{\partial l_2} = 0$

अथवा $2l_2 \sigma_2^2 - 2l_1 \sigma_1 \sigma_2 \rho = \lambda \quad \dots\dots(C)$

इन दोनों समीकरणों का हल ही हमारे प्रश्न का भी हल है। इनके अनुसार

$$\sigma_1 (l_1 \sigma_1 + l_2 \sigma_2 \rho) = \sigma_2 (l_2 \sigma_2 + l_1 \sigma_1 \rho)$$

अथवा $l_1 (\sigma_1^2 - \sigma_1 \sigma_2 \rho) = l_2 (\sigma_2^2 - \sigma_1 \sigma_2 \rho)$

$$\text{परन्तु } l_1 + l_2 = 1$$

$$\therefore l_1 = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2} \quad (B)$$

$$\text{और } l_2 = \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2} \quad (C)$$

इसी प्रकार यदि हमें एक ही प्राचल के अनेक प्राक्कलक ज्ञात हों तो हम उनका एक ऐसा एकघाती फलन मालूम कर सकते हैं जिसका प्रसरण न्यूनतम हो। इस प्रकार इन प्राक्कलकों के समस्त एकघाती फलनों में से वही सबसे अधिक दस होगा।

§ १७३ प्राक्कलक प्राप्त करने की कुछ विधियाँ

ऊपर दी हुई परिभाषाओं से आपको यह प्रतीत हुआ होगा कि किसी भी प्राचल के लिए पर्याप्त प्राक्कलक को खोज करनी चाहिए क्योंकि उसके द्वारा प्राचल के बारे में महत्तम सूचना हमें प्राप्त हो सकती है। परन्तु यह हमेशा सम्भव नहीं है। कई घटनों के लिए और कई प्राचलों के लिए कोई भी प्राक्कलक पर्याप्त नहीं है। इस कारण हमें दूसरी विधियाँ अपनानी पड़ती हैं। इनमें से कुछ जा विशेष महत्त्वपूर्ण हैं नीचे दी हुई हैं।

§ १७३१ महत्तम सभाविता विधि (maximum likelihood method)

मान लीजिए कि समष्टि असतत है और उसमें से एक यादृच्छिक प्रतिदर्श (x_1, x_2, \dots, x_n) का घनन किया जाता है। θ इस समष्टि का एक प्राचल है। इस विशेष प्रतिदर्श के लिए सभाविता फलन L को निम्नलिखित समीकरण द्वारा परिभाषित किया जाता है

$$L(x_1, x_2, \dots, x_n, \theta) = p_1(\theta) p_2(\theta) \dots p_n(\theta) \quad (176)$$

जहाँ $p_i(\theta)$ x_i के एक ऐसी समष्टि से चुन जाने की प्रायिकता है जिसका प्राचल θ हो।

यदि घटन सतत हो तो ऊपर लिखे ढग से सभाविता फलन की परिभाषा देना व्यथ है क्योंकि इस स्थिति में प्रत्येक x_i के लिए $p_i(\theta) = 0$ । इसलिए सतत घटनों के लिए प्रतिदर्श के सभाविता फलन को निम्नलिखित रूप में रल सकते हैं।

$$L(x_1, x_2, \dots, x_n, \theta) = f(x_1, \theta) f(x_2, \theta) \dots f(x_n, \theta) \quad (177)$$

जहाँ $f(x, \theta)$ θ प्राचल वाली समष्टि का x पर प्रायिकता घनत्वफलन है। उस θ का पता चलाने को जिसके लिए प्रतिदर्श का सभाविता फलन महत्तम हो जाय,

महत्तम सभाविता विधि कहते हैं। इस मान $\hat{\theta}$ का θ के प्राक्कलक की तरह उपयोग किया जाता है।

क्योंकि L घनात्मक है इसलिए $\log L$ का भी परिकलन किया जा सकता है। यह L का एक ऐसा फलन है जो L के साथ बढ़ता है। इसलिए θ के जिस मान के लिए L महत्तम है उसके लिए $\log L$ भी महत्तम है। $\log L$ का महत्तम मान मालूम करने के लिए हमें निम्नलिखित समीकरण हल करना पड़ेगा।

$$\left. \frac{\partial \log L}{\partial \theta} \right]_{\theta = \hat{\theta}} = 0 \quad (178)$$

इस समीकरण के हल को हम θ का महत्तम सभाविता प्राक्कलक (maximum likelihood estimator) कहते हैं। इन प्रकार के प्राक्कलन के कुछ गुण हैं जिनके कारण इसका विशेष महत्त्व है।

(१) यदि θ का कोई दक्ष प्राक्कलक $\hat{\theta}$ है तो सभाविता समीकरण का केवल एक हल होगा और वह होगा $\hat{\theta}$ । इस प्रकार यदि कोई दक्ष प्राक्कलक विद्यमान है तो इस विधि से उसका पता चल जाता है।

(२) यदि θ या कोई पर्याप्त प्राक्कलक $\hat{\theta}$ है तो सभाविता समीकरण का हल $\hat{\theta}$ का फलन होगा।

(३) कुछ प्रतिबंध ऐसे होते हैं, जो प्रायः सभी सम्बन्धियों द्वारा सन्तुष्ट हो जाते हैं। इनके अन्तर्गत सभाविता समीकरण का हल सगत् होता है।

(४) यह तो स्पष्ट ही है कि सभाविता समीकरण प्रेक्षित प्रतिदर्श पर आधारित है। इसलिए इसका हल एक यादृच्छिक चर है। बड़े प्रतिदर्शों के लिए इसके हल का वचन प्रायः प्रसामान्य होता है।

(५) बड़े प्रतिदर्शों के लिए यह हल प्रायः दक्ष होता है। यदि $\hat{\theta}_n$ एक महत्तम सभाविता प्राक्कलक है और $\hat{\theta}_n^*$ एक अन्य प्राक्कलक है तो हम एक ऐसी संख्या N मालूम कर सकते हैं कि यदि $n > N$ तो $V(\hat{\theta}_n) \leq V(\hat{\theta}_n^*)$

आइए, अब हम कुछ प्राचलों के प्राक्कलन के लिए इस विधि का प्रयोग करके देखें।

(I) समष्टि में केवल दो मान हैं 0 और 1 जिनकी प्रायिकता क्रमशः $1-p$ और p है। हम n परिमाण का एक प्रतिदर्श लेते हैं जिसमें r मान 1 और बाकी $(n-r)$ शून्य हैं। इस प्रतिदर्श के आधार पर p का प्राक्कलन करना है।

$$L = p^r (1-p)^{n-r}$$

$$\log L = r \log p + (n-r) \log (1-p)$$

$$\frac{\partial \log L}{\partial p} = \frac{r}{p} - \frac{n-r}{1-p}$$

इसलिए सभावितता समीकरण निम्नलिखित है

$$\frac{r}{\hat{p}} - \frac{n-r}{1-\hat{p}} = 0$$

$$\text{अथवा } r(1-\hat{p}) - (n-r)\hat{p} = 0$$

$$\text{अथवा } \hat{p} = \frac{r}{n}$$

(II) समष्टि प्वासो है जिसका प्राचल λ है। हम प्रतिदर्श x_1, x_2, \dots, x_n द्वारा λ का प्राक्कलन करना चाहते हैं।

$$L = e^{-\lambda} \frac{\lambda^{x_1}}{x_1!} \times e^{-\lambda} \frac{\lambda^{x_2}}{x_2!} \times \dots \times e^{-\lambda} \frac{\lambda^{x_n}}{x_n!}$$

$$= e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{x_1! x_2! \dots x_n!}$$

$$\log L = -n\lambda + \left(\sum_{i=1}^n x_i \right) \log \lambda - \log (x_1! x_2! \dots x_n!)$$

सभावितता समीकरण निम्नलिखित होगा

$$\left. \frac{\partial \log L}{\partial \lambda} \right]_{\lambda=\hat{\lambda}} = 0$$

$$\text{अथवा } -n + \frac{\sum_{i=1}^n x_i}{\hat{\lambda}} = 0$$

$$\therefore \hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$$

(III) यदि समष्टि $N(\mu, \sigma)$ हो तो

$$L(x_1, x_2, \dots, x_n, \mu, \sigma) = \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2}$$

$$\log L = -\frac{n}{2} \log(2\pi) - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

$\hat{\mu}$ के लिए सन्नहिता समीकरण निम्नलिखित है

$$\left. \frac{\partial \log L}{\partial \mu} \right]_{\mu=\hat{\mu}, \hat{\sigma}=\hat{\sigma}} = 0$$

$$\text{अथवा } \frac{\sum_{i=1}^n (x_i - \hat{\mu})}{\hat{\sigma}^2} = 0$$

$$\text{अथवा } \hat{\mu} = \bar{x}$$

$\hat{\sigma}$ के लिए सन्नहिता समीकरण निम्नलिखित है

$$\left. \frac{\partial \log L}{\partial \sigma} \right]_{\mu=\hat{\mu}, \sigma=\hat{\sigma}} = 0$$

$$\text{अथवा } -\frac{n}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^3} \sum_{i=1}^n (x_i - \hat{\mu})^2 = 0$$

$$\text{अथवा } \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2$$

$$\text{परंतु } \hat{\mu} = \bar{x}$$

$$\therefore \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

इस अंतिम उदाहरण में हम देखते हैं कि यदि समष्टि में दो या अधिक अज्ञात प्राचल हों तो उन्हें युगपत् (simultaneous) सन्नहिता समीकरणों की सहायता से प्राक्कलित किया जा सकता है।

यदि μ ज्ञात होता और केवल σ^2 का प्राक्कलन करना होता तो महत्तम सभाविता प्राक्कलक निम्नलिखित होता

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

यह देखा जा सकता है कि महत्तम सभाविता प्राक्कलक हमेशा अनभिन्नत नहीं होता । उदाहरण के लिए

$$\begin{aligned} E(\hat{\sigma}^2) &= \frac{1}{n} E \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \frac{1}{n} \sum_{i=1}^n E(x_i^2) - E(\bar{x}^2) \\ &= \frac{1}{n} \{n(\sigma^2 + \mu^2)\} - \mu^2 + \frac{\sigma^2}{n} \\ &= \sigma^2 \left(1 - \frac{1}{n}\right) \neq \sigma^2 \end{aligned}$$

§ १७.३.२ घूर्ण-विधि (method of moments)

किसी समष्टि के घूर्ण उसके प्राचलो के फलन होते हैं । यदि किसी समष्टि के k प्राचल $\theta_1, \theta_2, \dots, \theta_k$ हैं तो हम निम्नलिखित समीकरणों द्वारा इन प्राचलों के प्राक्कलनों को प्राप्त करते हैं

$$m'_i = \mu'_i \quad i=1, 2, \dots, k$$

जहाँ m'_i प्रतिदर्श का i -वाँ और μ'_i समष्टि का i -वाँ शून्यांतरिक घूर्ण (raw moment) है । (देखिए अध्याय २)

यह सिद्ध किया जा सकता है कि जिन प्रतिवधों को प्राय मभी समष्टियाँ सतुष्ट कर देती हैं उनके अतन्त इस प्रकार के प्राक्कलकों का बटन बड़े प्रतिदर्श परिमाणों के लिए प्राय प्रसामान्य होता है । यह प्राक्कलक सगत भी होते हैं, परन्तु हमेशा अनभिन्नत नहीं होते । बड़े प्रतिदर्शों के लिए यह प्राय दक्ष भी नहीं होते ।

प्राचल और प्रसामान्य बटनों के लिए तो यह विधि बहुत ही सरल है क्योंकि प्राचल स्वयं समष्टि के घूर्ण होते हैं । आइए, अब हम एक ऐसी समष्टि और ऐसे प्राचल का उदाहरण लें जिसके लिए प्राचल समष्टि का कोई घूर्ण नहीं होता हो ।

मान लीजिए यह समष्टि निम्नलिखित है ।

$$f(x, \lambda) = \frac{\alpha^\lambda}{\Gamma(\lambda)} e^{-x} x^{\lambda-1} \quad \left. \begin{array}{l} \alpha > 0 \\ 0 < x < \infty \end{array} \right\}$$

जिसमें λ एक ज्ञात अक्षर है।

$$\begin{aligned} E(x) &= \int_{-\infty}^{\infty} \frac{\alpha^\lambda}{\Gamma(\lambda)} x^\lambda e^{-x} dx \\ &= \frac{\alpha^\lambda}{\Gamma(\lambda)} \frac{\Gamma(\lambda+1)}{\alpha^{\lambda+1}} \\ &= \frac{\lambda}{\alpha} \end{aligned}$$

∴ α के प्रायकलक α^* के लिए निम्नलिखित समीकरण है

$$\frac{\lambda}{\alpha^*} = \frac{\lambda}{\alpha}$$

अथवा $\alpha^* = \frac{\lambda}{\bar{x}}$

इसी प्रकार घूर्ण विधि से प्राचलो का प्रायकलन बहुधा अत्यंत सरल हो जाता है।

§ १७४ विश्वास्य अंतराल (Confidence interval)

जो फलन प्रतिदर्श के लिए एक अद्वितीय मान ग्रहण करता हो उसके द्वारा 0 का प्रायकलन करने के स्थान में हम एक ऐसे अंतराल का भी प्रायकलन कर सकते हैं जिसमें 0 के होने की प्रायिकता एक पूर्व-निश्चित सख्या हो। पहिले तरीके को बिंदु-प्रायकलन (point estimation) और दूसरे तरीके को अंतराल प्रायकलन (interval estimation) कहते हैं।

मान लीजिए, प्रतिदर्श x_1, x_2, \dots, x_n ऐसी समष्टि से चुना गया है जिसको केवल एक प्राचल θ द्वारा निर्धारित किया जा सकता है। यदि t एक ऐसा प्रतिदर्शज है जो x_1, x_2, \dots, x_n तथा θ का फलन है परंतु जिसका बटन θ से स्वतंत्र है तो हम एक मान t_1 ऐसा मालूम कर सकते हैं कि t के इससे छोटे होने की प्रायिकता एक पूर्व-निश्चित सख्या α हो जहाँ $0 < \alpha < 1$ ।

अर्थात् $P[t \leq t_1] = \alpha$

यह संभव है कि असमता $t \leq t_1$ को हम एक दूसरे रूप $\theta \leq t_1^\alpha$ अथवा $\theta \geq t_1^\alpha$ में रख सकें। उदाहरण के लिए यदि समष्टि $N(\mu, 1)$ हो तो $t = (\bar{x} - \mu)$ एक ऐसा प्रतिदर्शज है जो x_1, x_2, \dots, x_n और μ का फलन है परंतु $(\bar{x} - \mu)$ का वटन $N\left(0, \frac{1}{\sqrt{n}}\right)$ है जो μ से स्वतंत्र है।

$$P\left[t \leq \frac{1.96}{\sqrt{n}}\right] = 0.975$$

$$P\left[\bar{x} - \mu \leq \frac{1.96}{\sqrt{n}}\right] = 0.975$$

अथवा $P\left[\mu \geq \bar{x} - \frac{1.96}{\sqrt{n}}\right] = 0.975$

(देखिए सारणी सख्या 8.2)

साधारणतया हम ऐसे दो मान t_1^α और t_2^α मालूम करना चाहते हैं कि

$$P\left[t_1^\alpha \leq \theta \leq t_2^\alpha\right] = \alpha \quad (17.10)$$

अंतराल (t_1^α, t_2^α) को हम θ का विश्वास-अंतराल (confidence interval) कहते हैं। जिसका विश्वास गुणांक (confidence coefficient) α है। ऊपर के उदाहरण में।

$$\begin{aligned} & P\left[\bar{x} - \frac{1.96}{\sqrt{n}} \leq \mu \leq \bar{x} + \frac{1.96}{\sqrt{n}}\right] \\ &= 1 - P\left[\bar{x} > \mu + \frac{1.96}{\sqrt{n}}\right] - P\left[\bar{x} < \mu - \frac{1.96}{\sqrt{n}}\right] \\ &= 1 - P\left[(\bar{x} - \mu)\sqrt{n} > 1.96\right] - P\left[(\bar{x} - \mu)\sqrt{n} < -1.96\right] \\ &= 1 - 0.025 - 0.025 \\ &= 0.95 \end{aligned}$$

मान लीजिए किसी प्रतिदर्श के लिए $\bar{x} = 10$ $n = 4$ क्या हम कह सकते हैं कि

$$P[9.02 \leq \mu \leq 10.98] = 0.95$$

इस तरह का वक्तव्य देना अर्थहीन होगा क्योंकि प्रायिकता वक्तव्य किसी यादृच्छिक चर अथवा यादृच्छिक घटना के संबन्ध में ही दिये जा सकते हैं और ऊपर के वक्तव्य में इस प्रकार की किसी यादृच्छिक घटना की कल्पना नहीं की गयी है ।

$$P\left[\bar{x} - \frac{1.96}{\sqrt{n}} \leq \mu \leq \bar{x} + \frac{1.96}{\sqrt{n}}\right] = 0.95 \text{ एक अर्थपूर्ण वक्तव्य है}$$

क्योंकि $\left(\bar{x} - \frac{1.96}{\sqrt{n}}, \bar{x} + \frac{1.96}{\sqrt{n}}\right)$ एक यादृच्छिक अंतराल है जिसमें μ के पाये जाने की प्रायिकता का कुछ अर्थ है । यदि हम बार-बार इस समष्टि में से n परिमाण के प्रतिदर्श लें और इन अंतराल का प्राक्कलन ऊपर दिये हुए सूत्र द्वारा करें तो हम आशा कर सकते हैं कि 95 प्रतिशत अंतराल ऐसे होंगे जिनमें μ पाया जायगा । और केवल 5 प्रतिशत अंतराल ही ऐसे होंगे कि μ उनके बाहर हो ।

क्योंकि हमारा प्रतिदर्श इस समष्टि में से चुना गया है और क्योंकि अंतराल का प्राक्कलन हम विशेष विधि से किया गया है, इसलिए हमें विश्वास है कि μ इस अंतराल में ही होगा । यदि अंतराल इस प्रकार के अंतरालों में से चुना जाता जिनमें से 90 प्रतिशत में ही μ पाया जाता तो भी हमें यह विश्वास होता कि θ उसी के अंतर्गत है । परन्तु इस विश्वास की मात्रा अपेक्षाकृत कम होती । किसी अपनायी हुई विधि से प्राप्यकलित अंतरालों में θ के पाये जाने की प्रायिकता को हम इस विश्वास की मात्रा का माप मान सकते हैं । इसी कारण इसको विश्वास-गुणांक कहा जाता है ।

भाग ५

प्रयोग अभिकल्पना

Design of Experiment

अध्याय १८

संपरीक्षण (experimentation) में सांख्यिकी का स्थान

§ १८१ भौतिकी और रसायन के प्रयोगों में सांख्यिकी का साधारण-सा महत्त्व

विज्ञान का इतिहास प्रयोगों (experiments) और उनके फलों को समझने के प्रयत्नों का इतिहास है। विज्ञान की अन्य शाखाओं की अपेक्षा भौतिक और रसायन अधिक पुरातन है। इनमें प्रयोगों की विधि इतनी उन्नत हो चुकी है कि साधारणतया प्रयोगों के फलों में कोई विशेष अंतर नहीं पड़ता, चाहे उन्हें कोई भी व्यक्ति किसी भी स्थान पर और किसी भी समय क्यों न करे। यदि कुछ विशेष अंतर लाया भी जाये तो उसकी व्याख्या तापमान, वायुदाब आदि गिने चुने उपादानों (factors) द्वारा हो सकती है। ऐसे समीकरण ढूँढ निकाले गये हैं जो प्रयोगों के फलों को इन उपादानों के फलन के रूप में व्यक्त कर सकते हैं। यह सच है कि प्रयोग के फल और इस फलन के मान में फिर भी कुछ अंतर रह ही जाता है। परंतु यह अंतर इतना कम होता है कि इसे प्रायोगिक त्रुटि (observational error) समझ लिया जाता है। इस प्रकार के विज्ञान में अथवा उनके विकास के लिए किये गये प्रयोगों में सांख्यिकी का कोई स्थान नहीं है। हाँ, इसमें गाउस (Gauss) के त्रुटि-वटन का प्रयोग यदा-कदा कर लिया जाता है। इसके अलावा सांख्यिकी के इस सिद्धांत का प्रयोग भी बहुधा किया जाता है कि प्रतिदर्श-परिमाण बढ़ने के साथ साथ प्रतिदर्श माध्य का प्रसरण कम होता जाता है। इसी कारण विज्ञान में यह प्रथा है कि एक ही माप में प्रयोगकर्ता सतुष्ट नहीं होता। वह एक ही प्रयोग के फलों का भी अनेकों बार माप लेता है। प्रयोगों के फलों का विभिन्न उपादानों से सबसे स्थापित करने के लिए समीकरण में इन मापों के माध्य का ही प्रयोग किया जाता है।

§ १८२ विज्ञान की अन्य शाखाओं में सांख्यिकी का असाधारण महत्त्व यद्यपि विज्ञान की इन महत्त्वपूर्ण शाखाओं में सांख्यिकी का कोई विशेष स्थान नहीं है, परंतु अन्य विभागों में विशेषकर प्राणि-विज्ञान और सामाजिक विज्ञान में

सांख्यिकी ने अपने लिए बहुत महत्त्वपूर्ण स्थान बना लिया है। इन विज्ञानों में नियम अधिकतर यथार्थ न होकर सांख्यिकीय होते हैं। परन्तु यहाँ हमें इन विज्ञानों के नियमों अथवा सिद्धांतों में कोई दिलचस्पी नहीं है। हम तो यह देखना चाहते हैं कि स्वयं सपरीक्षण अथवा प्रयोग-विधि (experimentation) को सांख्यिकी ने कहाँ तक प्रभावित किया है। साधारणतया सांख्यिक स्वयं कोई वैज्ञानिक प्रयोग नहीं करते, परन्तु फिर भी पिछले कुछ वर्षों में सांख्यिकों द्वारा सपरीक्षण विधि पर कई लेख व पुस्तकें लिखी जा चुकी हैं। यह माना जाने लगा है कि वैज्ञानिकों को, जो प्रयोग करके उनके फलों का समुचित उपयोग करना चाहते हैं, इन सांख्यिकीय साहित्य से किसी हद तक परिचित होना आवश्यक है। यदि वे इससे परिचित नहीं हैं या उन्हें किसी विशेष परिस्थिति का सामना करना है तो उन्हें सांख्यिकों से सलाह लेनी चाहिए। अनुसंधानकर्ता प्रयोग-विधि निश्चित करने में और प्रयोग के फलों की व्याख्या करने में सांख्यिकी और सांख्यिका का सहारा इतना अधिक लेने लगे हैं कि कुछ वैज्ञानिकों की राय में अब यह सहारा उचित सीमा का उल्लंघन कर चुका है और वे उसके ऊपर रोक लगाना चाहते हैं। यद्यपि हम इन कतिपय वैज्ञानिकों से सहमत हैं कि कदाचित् सांख्यिकी का आवश्यकतासे अधिक और अनुचित प्रयोग होने लगा है, परन्तु प्रयोग अभिकल्पना (design of experiments) में सांख्यिकी ने जो स्थान बना लिया है उससे अब उसे हटा देना अनभव है।

§ १८.३ परिकल्पना की जांच और प्राचलो के प्राक्कलन में प्रयोग अभिकल्पना का महत्त्व

यह हम पहले ही कह चुके हैं कि भौतिक और रसायन के प्रयोगों के फलों के विपरीत अन्य विज्ञानों में प्रयोग को बार-बार दुहराने पर उसके फल भिन्न भिन्न होते हैं। यह हो सकता है कि यदि उन सभी उपादानों को स्थिर रखा जाय जो प्रयोग पर प्रभाव डालते हैं तो इन फलों में भी अंतर न आये। लेकिन अभी तक न तो वैज्ञानिकों को इन सब उपादानों का ज्ञान है और न ही वे ज्ञात उपादानों को नियंत्रित करने की कठिनाइयों पर विजय प्राप्त कर पाये हैं। यही नहीं, बल्कि इनका विरवास है कि सब छोटे-छोटे उपादानों के प्रभाव का ज्ञान बहुत महत्त्वपूर्ण नहीं होता। अधिक महत्त्वपूर्ण तो यह जानना है कि इन उपादानों का सचित प्रभाव क्या है। कुछ भी हो, यह सच है कि इन प्रयोगों की प्रकृति यादृच्छिक प्रयोगों की सी ही होती है जिनका वर्णन पहिले ही कई बार किया जा चुका है।

हम पिछले कुछ अध्यायों में यह देख ही चुके हैं कि प्रयोग के फलों की व्याख्या किसी हद तक परिकल्पना की जाँच द्वारा किस प्रकार की जा सकती है। इसी प्रकार हम यह भी देख चुके हैं कि प्रतिदर्श से समष्टि के प्राचलों (parameters) का प्राक्कलन (estimation) किस प्रकार किया जाता है। किन्तु अभी तक हमने इस समस्या पर भली भाँति विचार नहीं किया है कि प्रयोग किस प्रकार किये जाएँ अथवा प्रतिदर्श किस प्रकार चुने जायँ कि उनको वास्तव में यादृच्छिक की सजा दी जा सके और उनके फलों को यादृच्छिक चर समझा जाना युक्तियुक्त हो। इन यादृच्छिक चरों के प्रायिकता-वटन का ज्ञात होना ही प्रयोग की व्याख्या को संभव बनाता है। यदि ऐसा हम नहीं कर पाये तो कुछ थोड़े से प्रयोगों के फलों से अथवा एक प्रतिदर्श से प्राचलों का अनुमान लगाना बहुत कठिन हो जायगा।

§ १८.४ उदाहरण

मान लीजिए, एक रोग के लिए दो औषधों की तुलना हम करना चाहते हैं। यदि इन औषधों का सौ-सौ रोगियों पर प्रयोग किया जाय तो हम जानते हैं कि परिकल्पना क्या होनी चाहिए और उसकी जाँच कैसे करनी होगी। परन्तु इस जाँच के लिए द्विपद-वटन अथवा प्रसामान्य-वटन का उपयोग हम उसी दशा में कर सकते हैं जब इन रोगियों को संपूर्ण रोगी-जगत् का प्रतिनिधि मान लेना किसी हद तक मुक्तिसंगत हो। यदि इन रोगियों का चुनाव यादृच्छिक हो तब तो इन वटनों का उपयोग संगत है ही—कुछ अन्य परिस्थितियों में भी इसे ठीक समझा जा सकता है।

परन्तु अनेक प्रयोग इस प्रकार किये जाते हैं कि उनसे कोई लाभदायक अनुमान लगाना मुश्किल है। उदाहरण के लिए यदि सभी रोगियों पर एक ही औषध का प्रयोग किया जाय तो उसके उपयोग को नहीं मालूम किया जा सकता। अथवा यदि सभी रोगी जिन्हें एक विशेष औषध दी जाय, एक विशेष अस्पताल के हों तथा अन्य रोगी जिन्हें दूसरी औषध दी जाय दूसरे अस्पताल के हों तो यह प्रयोग गलत होगा। रोगी के नीरोग होने का कारण बेचल औषध नहीं होती। उसका भोजन, धाराम और सफाई आदि का प्रबंध भी उसके नीरोग होने की प्रायिकता को प्रभावित करते हैं। इस दशा में यदि किसी अस्पताल के रोगियों में से एक बहुत बड़ा अनुपात नीरोग हो जाता है जब कि दूसरे में केवल थोड़े से रोगी स्वास्थ्य लाभ कर पाते हैं तो यह कैसे कहा जा सकता है कि यह अंतर औषधों के प्रभाव के कारण है अथवा दोनों अस्पतालों में रोगियों की देखभाल के भेद के कारण। इस प्रकार किसी प्रयोग का फल यदि कई

उपादानों पर निर्भर हो और हम उनमें से केवल एक का प्रभाव जानना चाहते हैं तो अन्य उपादानों के प्रभाव से छुटकारा पाना आवश्यक हो जाता है।

ऊपर के उदाहरण में दोनों औषधों का प्रभाव जानने के लिए यदि दोनों अस्पतालों से पचास-पचास रोगियों के प्रतिदर्श लिये जायें तो अस्पताल के प्रभावों में छुटकारा पाया जा सकता है। परन्तु रोगी के नीरोग होने की प्रायिकता उसकी उम्र और साधारण स्वास्थ्य पर भी तो निर्भर करती है। यदि भूल से हमारे प्रतिदर्श में एक औषध के लिए अधिकतर रोगी बूढ़ और निबल हो और जिन रोगियों को दूसरी औषध दी जाय उनमें अधिकतर जवान तथा हृष्टपुष्ट हो तो भी औषध के बारे में अनुमान लगाना कठिन है। हो सकता है कि इन उपादानों के प्रभाव को हटाने के लिए आप प्रतिदर्श का चुनाव इस प्रकार करें कि उम्र का वितरण दोनों प्रतिदर्शों में समान हो। लेकिन किसी रोगी के नीरोग होने अथवा मृत्यु-लाभ करने में इतने अधिक उपादानों का प्रभाव पड़ता है कि उन सबके प्रभावों को बिल्कुल हटा देना असंभव है। कुछ तो यह इस कारण है कि सब उपादान शत नहीं हैं और कुछ इस कारण कि ज्ञात उपादानों की संख्या भी इतनी अधिक है कि उनका नियंत्रण करने के लिए भी बहुत बड़े प्रतिदर्श की आवश्यकता होगी। इतने बड़े प्रतिदर्श पर प्रयोग करने के लिए खर्चा भी बहुत अधिक होगा और यह संभव है कि उतना रुपया उपलब्ध ही न हो। और यदि हो भी तो शायद इतने अधिक रोगियों को प्रयोग के लिए ढूँढना मुश्किल हो। यदि रोगी भी मिल जायें तो भी इतने बड़े प्रयोग को भली भाँति नियंत्रित करने में अनेक कठिनाइयाँ हैं। यह देखना कि रोगियों को ठीक समय पर औषध दी जा रही है अथवा नहीं, उनके भोजन और आराम आदि की व्यवस्था ठीक है अथवा नहीं, उनका प्रेक्षण करने के लिए प्रशिक्षित प्रेक्षकों (observers) को पर्याप्त संख्या में प्राप्त करना आदि अनेक कठिनाइयाँ हैं।

§ १८५ यादृच्छिकीकरण (Randomization)

यदि प्रयोग छोटे पैमाने पर हो तो उसका नियंत्रण कठोरता से हो सकता है। यदि छोटे पैमाने के इस प्रयोग से भी समाष्ट के बारे में अनुमान लगाना संभव हो तो हम व्यर्थ में प्रयोग को बढ़ाकर अधिक खर्च के साथ-साथ अन्य कठिन समस्याओं को क्यों निमग्नित करें? यह स्पष्ट है कि इस छोटे-से प्रयोग द्वारा हम सब उपादानों के प्रभाव को पूरी तौर से हटा नहीं सकते, परन्तु इनके कारण प्रयोग में जो अभिन्नता (bias) आ सकती है उससे बचने के लिए एक तरीका है।

इस तरीक़े का नाम है “यादृच्छिकीकरण” (*randomization*) जिसका आविष्कार प्रोफ़ेसर रोनाल्ड ए० फ़िशर ने किया था । इसके अनुसार कौन-सी औषध किन रोगियों को दी जायगी, यह एक यादृच्छिक प्रयोग द्वारा निश्चित किया जाता है । उदाहरण के लिए हर एक रोगी के लिए एक सिक्का उछालकर निश्चित किया जा सकता है कि उसे पहली औषध दी जाय या दूसरी । इसका फल यह होता है कि दोनों औषधों को अधिक पढ़ अथवा अधिक हृष्ट-युष्ट रोगियों का इलाज करने का बराबर मौका मिलता है । यह ही सकता है कि किन्हीं विशेष यादृच्छिक प्रयोग के फलस्वरूप एक औषध के लिए परिस्थिति अनुकूल हो और दूसरी के लिए प्रतिकूल हो, क्योंकि रोगियों के दोनों समूह बिल्कुल एक समान तो हो सकते नहीं । लेकिन यह अंतर जितना होता है उसका विचार पहिले ही परिकल्पना की जाँच और विश्वास्य सीमाओं के परिकलन में कर लिया जाता है । प्रयोग की अभिकल्पना में ऐसी बहुत कम विशेषताएँ हैं जो वास्तव में आपुनिक हैं । इन कुछ विशेषताओं में यादृच्छिकीकरण एक है । यादृच्छिकीकरण का किस स्थान पर किस प्रकार उपयोग किया जाय यह बहुत कुछ प्रयोग करनेवाले की विवेक-बुद्धि पर निर्भर करता है । ऊपर के उदाहरण में यह काफी है कि कुल रोगियों में से आधे का यादृच्छिक चुनाव किया जाय जिनको पहली औषध देनी है और बाकी रोगियों को दूसरी दवा दे दी जाय । इस विधि में हर एक रोगी के लिए इन दो दवाओं द्वारा इलाज करवाये जाने की प्रायिकताओं को बराबर होना चाहिए । कई अन्य प्रयोगों में—उदाहरण के लिए मनोवैज्ञानिक प्रयोगों में—कई ऐसी क्रियाएँ होती हैं जो अभिनति का कारण हो सकती हैं । बहुधा जिन व्यक्तियों पर ये प्रयोग किये जाते हैं उनमें ही अन्तर पड़ जाता है । वे प्रयोग के दौरान में कुछ अधिक सीख जाते हैं अथवा एकान के कारण उनकी कार्य-क्षमता में अन्तर आ जाता है । ऐसी व्यवस्थित अभिनति से बचने के लिए यादृच्छिकीकरण का उपयोग किया जाता है । अन्य कठिन अवस्थाओं में यादृच्छिकीकरण का एक ही प्रयोग में बार-बार उपयोग करना पड़ सकता है ।

कई बार हमें विश्वास होता है कि बिना यादृच्छिकीकरण के कोई विशेष अभिनति नहीं होनी चाहिए । इस पर भी यह उचित है कि इस सांख्यिकीय क्रिया के करने का कष्ट उठाया जाय । इसके द्वारा प्रयोगकर्ता अनपेक्षित घटनाओं से प्रयोग के बेकार हो जाने की सम्भावना को दूर कर सकता है । किसी विशेष प्रयोग में इतनी अधिक क्रियाएँ हो सकती हैं कि उन सबके लिए यादृच्छिकीकरण में बहुत समय और धन व्यय होने की आशंका है और वदाचित् उससे इतना लाभ न हो । इस परिस्थिति

में प्रयोगकर्ता को निश्चय करना पड़ता है कि कौन-सी क्रियाएँ अभिनति के दृष्टिकोण से अधिक महत्त्वपूर्ण हैं और यादृच्छिकीकरण को केवल इन क्रियाओं तक ही सीमित रखना पड़ता है ।

§ १८.६ नियन्त्रित यादृच्छिकीकरण

यद्यपि इस यादृच्छिकीकरण से अभिनति का परिहार हम कर सकते हैं, फिर भी किसी औषध को विशेष सुविधा (advantage) मिलने की संभावना को पूर्णतया संयोग पर छोड़ना बुद्धिमानी नहीं है । कम से कम कुछ उपादानों के प्रभाव को दोनों औषधों के लिए बराबर-बराबर बाँटने की चेष्टा हमें अवश्य करनी चाहिए । जैसा कि हम पहिले विचार कर चुके हैं, दोनों अस्पतालों में बराबर-बराबर सख्या के रोगियों को उन दोनों प्रकार की औषधों का दिया जाना अधिक उचित जान पड़ता है । यदि हीं सके तो रोगियों के उन दोनों वर्गों में—जो इन दो दवाओं का सेवन करने के लिए चुने गये हों—उम्र का वटन और स्वास्थ्य की स्थिति एक समान कर देनी चाहिए । यद्यपि केवल इन्हीं दो उपादानों के प्रभाव से बचाना ही काफी नहीं है तथापि शायद कुल उपादानों के सम्पूर्ण प्रभाव का एक बहुत बड़ा भाग इन्हींके कारण है । हम पूर्ण विश्वास के साथ इनको नियन्त्रित करने का जिम्मा सिर्फ संयोग पर नहीं छोड़ सकते । इसके लिए हमें अन्य तरीके अपनाने होंगे । दूसरी ओर धारणें शायद यह भी सोचा हो कि परिकल्पना की जाँच के लिए आवश्यक है कि प्रयोग के फल यादृच्छिक घर हों और इस कारण यादृच्छिकीकरण का सर्वथा त्याग उचित नहीं है । ऐसा करने से सपूर्ण प्रयोग के वृथा हो जाने की संभावना है । ऐसी दशा में क्या करना चाहिए ? इस समस्या को मुलज्ञाने के लिए बहुत सांख्यिकीय ज्ञान की आवश्यकता नहीं है । यदि आप ध्यानपूर्वक इस पर विचार करें तो समस्या को मुलज्ञा सकते हैं । यद्यपि इस के कई हल हो सकते हैं, परन्तु उनमें से एक निम्नलिखित है ।

दो-दो रोगियों के अनेको युग्म (pairs) बनाये जा सकते हैं जिसमें दोनों रोगी जहाँ तक इन उपादानों का संबंध है, एक समान हों । यदि औषधियाँ A और B हों तो हमें इनमें से एक युग्म के लिए यह निर्णय करना होता है कि किस रोगी को A और किसको B दी जाय । यह एक यादृच्छिक प्रयोग द्वारा—उदाहरण के लिए एक सिक्के को उछालकर—किया जा सकता है । इस प्रकार हम इन उपादानों को नियन्त्रित भी कर लेते हैं और यादृच्छिकीकरण के उपयोग द्वारा अभिनति का परिहार भी हो जाता है । यदि दो न होकर औषधियों की सख्या n हो तो

हमें कुल रोगियों को ऐसे कुलका (sets) में बाँटना होगा जो कुछ महत्वपूर्ण उपादानों की दृष्टि से समाग हो और प्रत्येक कुलक में रोगियों की संख्या n हो ।

§ १८.७ ब्लॉक

प्रायोगिक इकाइयों के इन कुलकों को—जिनमें विभिन्न उपचारों (treatments) को इकाइयों में यादृच्छिकीकरण द्वारा बाँटा जाता है—सांख्यिकीय भाषा में ब्लॉक (block) कहते हैं। इसका कारण यह है कि प्रयोग की अभिकल्पना के सांख्यिकीय सिद्धांतों का भाविष्कार आरम्भ में कृपि सभी प्रयोगों के लिए ही किया गया था । उनमें यह कुलक एक सहत भूखंड (compact piece of land) होता है जिसे अंग्रेजी में अक्सर ब्लॉक भी कहते हैं । इसी प्रकार अन्य अनेक पारिभाषिक शब्द—जिनका प्रयोग-अभिकल्पना साहित्य में उपयोग होता है—कृपि से संबंधित हैं । परन्तु अब तक आप यह तो समझ ही चुके हैं कि इन सिद्धांतों का उपयोग कृपि-विज्ञान में ही नहीं बल्कि प्राणि-विज्ञान, मनोविज्ञान और सामाजिक-विज्ञान के प्रायः सभी प्रयोगों में होता है ।

§ १८.८ प्रयोग आरम्भ करने से पूर्व योजना की आवश्यकता

यह बहुधा देखा जाता है कि वैज्ञानिक प्रयोग के लिए योजना बनाते समय सांख्यिकी से सलाह लेने की आवश्यकता नहीं समझी जाती । जब वे प्रयोग कर चुकते हैं तो सकलित आँकड़ों को सांख्यिकी के सामने रखकर कहते हैं कि आप जरा इनका विश्लेषण और व्याख्या तो कर दीजिए । सांख्यिक प्रायः किसी विज्ञान में विशेष दक्ष नहीं होता और इसलिए उसे यह जानना आवश्यक हो जाता है कि प्रयोग किस उद्देश्य से किया गया था । इसके अलावा प्रयोग में जो विधि अपनायी गयी थी उसका जानना भी आवश्यक होता है । सांख्यिक चेष्टा करता है कि प्रयोग के उद्देश्य को किनी प्रकार सांख्यिकीय परिकल्पना के रूप में रख सके । फिर उसे यह देखना होता है कि प्रयोग के लिए जो विधि अपनायी गयी है उसके द्वारा इस परिकल्पना की जाँच होना कहीं तक संभव है ।

बुल उल्लाही जन प्रयोगों को बिना पूरी तरह योजना बनाये ही आरम्भ कर देते हैं । बाद में उन्हें यह मालूम होता है कि जिस प्रकार प्रयोग किया गया है उससे उद्देश्य-पूर्ति नहीं होती । अथवा प्रयोग में प्रतिदर्श परिमाण इतना कम था कि उसके आधार पर किनी निश्चित परिणाम पर पहुँचना संभव नहीं । कई बार प्रतिदर्श परिमाण इतना अधिक होता है कि उससे बहुत कम में ही काम चल सकता था । इन सब दशाओं

में प्रयोग में लगाये हुए धन और समय का अपव्यय होता है । यह कही अधिक अच्छा हो यदि सांख्यिक की सलाह योजना बनाते समय ही ले ली जाय । ऐसी अवस्था में वह यह आश्वासन दे सकती है कि प्रयोग के उद्देश्य में सफलता मिलने की संभावना है अथवा नहीं ।

§ १८ ९ प्रयोग की योजना बनाते समय तीन बातों का ध्यान रखना होता है

- (१) प्रयोग का उद्देश्य क्या है ?
- (२) प्रायोगिक इकाइयाँ क्या हैं ? प्रयोग किस प्रकार किया जा रहा है और प्रयोग में प्रतिदर्श-परिमाण क्या होगा ?
- (३) प्रायोगिक फल का विश्लेषण किस प्रकार किया जायगा ?

§ १८ १० प्रयोग का उद्देश्य

किमी भी प्रयोग का उद्देश्य एक या अधिक प्रतिदर्शों के आधार पर समष्टि के बारे में ज्ञान प्राप्त करना अथवा उससे संबंधित कुछ कथनों की सत्यता की जाँच करना होता है । सांख्यिक को यह मालूम होना चाहिए कि वह कौन-सी समष्टि है जिसके बारे में वैज्ञानिक ज्ञान प्राप्त करना चाहता है । मान लीजिए कि एक प्रयोग का उद्देश्य गेहूँ की फसल के लिए विभिन्न खादों के प्रभाव का पता लगाना है । परन्तु यह उद्देश्य सुस्पष्ट नहीं है । गेहूँ केवल एक ही प्रकार के नहीं होते । वे कई प्रकार के होते हैं । यह जानना आवश्यक है कि प्रयोगकर्ता किसी विशेष प्रकार के गेहूँ पर खादों के प्रभाव का अध्ययन करना चाहता है अथवा साधारणतया सभी प्रकार के गेहूँ पर । इसी प्रकार विभिन्न प्रदेशों के जलवायु और जमीन में अन्तर होता है । जो छद्म एक प्रदेश में लाभदायक सिद्ध होती है वह किसी दूसरे प्रदेश में बेकार भी हो सकती है । इस कारण यह जानना भी जरूरी है कि प्रयोगकर्ता की रुचि किसी प्रदेश विशेष में है अथवा साधारणतया सभी प्रदेशों में । इस प्रकार प्रयोग के फलों को प्रभावित करनेवाले उपादानों में से कौन ऐसे हैं जिन्हें स्थिर रखा जा सकता है, यह मालूम हो जाता है ।

यदि उद्देश्य बहुत महत्वाकांक्षायुक्त नहीं है—यदि किसी साधारण समष्टि के लिए किसी एक कथन की पुष्टि अथवा उसका खण्डन करना हो तो तुलनात्मक दृष्टि से काफी छोटे प्रतिदर्श को लेकर ही प्रयोग किया जा सकता है । यदि प्रयोगकर्ता बहुत महत्वाकांक्षी है तो संभव है कि उसकी आकांक्षा वर्षों प्रयोग करने पर भी पूरी न हो ।

समष्टि के बारे में फैसला हो जाने पर यह जानना आवश्यक है कि वह कथन क्या है जिसकी पुष्टि अथवा खंडन करना प्रयोग का उद्देश्य है। कुछ कथन ऐसे होते हैं जिनकी पुष्टि करना अथवा जिनका खंडन करना प्रयोगों द्वारा असंभव है। इस प्रकार के कथन अधिकतर महत्त्वहीन होते हैं। यदि वे महत्त्वपूर्ण हों भी तो बहुधा प्रयोगकर्ता अथवा सांख्यिक के पास उनकी जांच करने का कोई साधन नहीं होता।

ऊपर के उदाहरण के लिए कथन निम्नलिखित हो सकता है। “जाद A गेहूँ की फसल के लिए अन्य खादा की अपेक्षा अधिक अच्छी है।” प्रश्न यह उठता है कि यह किस दृष्टिकोण से अच्छी है? क्या उसके कारण गेहूँ की पैदावार अधिक होती है? क्या उसके कारण गेहूँ के पौधों में बीमारी से बचने की शक्ति बढ़ती है? क्या उसके कारण गेहूँ की पोष्टिकता (food value) बढ़ जाती है? क्या उसके कारण गेहूँ की फसल जल्दी तैयार हो जाती है? प्रयोग का उद्देश्य इनमें से एक या अधिक प्रश्नों का उत्तर प्राप्त करना हो सकता है, परंतु योजना के लिए इसका स्पष्ट-तया जानना आवश्यक है। इसके अलावा ये कथन इस प्रकार के होने चाहिए कि उन्हें एक सांख्यिकीय परिकल्पना के रूप में रखा जा सके।

§ १८११ प्रायोगिक उपचार (Experimental treatments)

उपचारों से हमारा तात्पर्य यहाँ उन विविध क्रियाओं से है जिनके प्रभाव को नापना और उनकी तुलना करना प्रयोग का उद्देश्य होता है। इन क्रियाओं की भली-भाँति व्याख्या करना आवश्यक होता है। हमें यह भी जानना चाहिए कि प्रयोग का उद्देश्य केवल सबसे प्रभावशाली साधन का पता चलाना है अथवा यह मालूम करना है कि इन साधनों के प्रेषित प्रभाव का कारण क्या है? यद्यपि कई व्यावहारिक समस्याओं को मुलझाने के लिए सर्वोत्तम साधन का जानना ही यथोष्ट होता है, परंतु कारण के ज्ञान से ही विज्ञान की उन्नति द्रुत गति से होती है। कई बार प्रयोग में हम कुछ ऐसे साधनों पर भी विचार करते हैं जिनके बारे में हम जानते हैं कि इनका व्यवहार कभी नहीं किया जायगा। इन साधनों का उपयोग प्रयोग में केवल कारण जानने के लिए किया जाता है।

§ १८१२ बहु-उपादानिय प्रयोग (Factorial experiments)

हम पहिले ही कह चुके हैं कि हमें यह जानना आवश्यक है कि किस उपादान के प्रभाव को हम नापना चाहते हैं। दूसरे उपादानों के प्रभाव को हम स्थिर रख सकते हैं। परंतु यह तभी ठीक होगा जब इन उपादानों के प्रभाव समोज्य (additive)

हों। यदि ऐसा हो तो यह निश्चित करने में कुछ भी कठिनाई नहीं पड़ती कि अन्य उपादानों को किस मान पर स्थिर रखा जाय। परन्तु यदि यह प्रभाव सयोग्य नहीं है तो किसी विशेष उपादान का प्रभाव उन मानों पर भी निर्भर हो सकता है जिन पर अन्य उपादानों को अन्तर रखा जाता है। ऐसी स्थिति में इस विशेष उपादान के प्रभाव को अन्य उपादानों के क्रम से कम दो विभिन्न मानों पर नापना ठीक समझा जाता है। इस प्रकार के प्रयोग में हम न केवल इस विशिष्ट अवयव या उपादान के बल्कि अन्य उपादानों के प्रभाव को भी नाप सकते हैं। इस प्रकार के प्रयोगों को बहु-उपादानिय प्रयोग (*factorial experiments*) कहा जाता है। आगे चलकर हम इन प्रयोगों की विधि और उनके विश्लेषण पर विस्तारपूर्वक विचार करेंगे।

§ १८.१३ नियंत्रण इकाइयाँ (*Control units*)

कई बार ऐसा होता है कि जिन इकाइयों पर प्रयोग किया जाता है उनकी किसी विशेषता के कारण प्रयोग ध्येय हो जाता है। उदाहरण के लिए एलोपैथी और होमियोपैथी की तुलना को ही लीजिए। आपको शायद पता होगा कि कई शारीरिक रोग केवल मनोदशाजनित अथवा मन शारीरिक (*psychosomatic*) होते हैं। उनका कारण कोई भौतिक पदार्थ, रसायन, विष अथवा कीटाणु नहीं होता। यदि रोगी को किसी वजह से यह ब्याल हो जाय कि उसका स्वास्थ्य ठीक नहीं है तो उसको यह मनोदशा ही रोग का कारण बन सकती है। यदि रोगी को पता न लगे और वह यह समझे कि उसे कोई बहुत गुणकारी औषध दी जा रही है तो केवल आटे की गोलियों अथवा शुद्ध जल से भी उसका इलाज हो सकता है। ऐसे रोगियों का यदि एलोपैथी अथवा होमियोपैथी द्वारा उपचार किया जाय तो उसका फल इस पर निर्भर करेगा कि रोगी को इनमें से किस पर विश्वास है। आरम्भ में यह पता लगाना कठिन है कि रोगियों में से वे कौन से हैं जिनका रोग मन शारीरिक है। ऐसी दशा में यद्यपि हमारा उद्देश्य केवल होमियोपैथी और एलोपैथी की तुलना करना है, तथापि हमें यह आवश्यक हो जाता है कि कुछ रोगियों पर इन दोनों में से किसी भी इलाज का प्रयोग नहीं किया जाय, बल्कि आटे की गोलियों जैसी निरर्थक दवाई इस्तेमाल की जाय। इस प्रयोग से हम मन शारीरिक रोग से पीड़ित रोगियों के अनुपात का अंदाजा लगा सकते हैं। इस प्रकार एक निरर्थक उपचार के प्रयोग से प्रयोग निरर्थक न रहकर सार्थक हो जाता है। इस प्रकार की इकाइयों को—जिनपर निरर्थक उपचार किया जाता है—नियंत्रण इकाइयाँ (*control units*) कहते हैं।

§ १८१४ प्रयोग-अभिकल्पना का एक सरल उदाहरण

यद्यपि वैज्ञानिक अनेक वर्षों से प्रयोग करते आ रहे हैं, परंतु उनकी अभिकल्पना और विवरण शैली को पहली बार व्यवस्थित रूप में रखने का श्रेय है प्रो० रोनाल्ड ए० फिशर को। अपनी (Design of Experiments) नाम की पुस्तक में उन्होंने अभिकल्पना के सिद्धांतों से परिचित होने के लिए एक कल्पित, परंतु बहुत ही दिलचस्प प्रयोग का उदाहरण दिया है। सांख्यिकीय साहित्य में यह उदाहरण बहुत प्रसिद्ध हो गया है और कुछ अन्य सांख्यिकी में भी इसी उदाहरण को लेकर प्रयोग-अभिकल्पना की व्याख्या की है। आगे इस कल्पित प्रयोग का संक्षेप में वर्णन किया गया है।

§ १८१४१ प्रयोग का उद्देश्य

एक महिला का यह दावा है कि वह चाय को चखकर यह बता सकती है कि प्याले में पहिले चाय डाली गयी थी अथवा दूध। हम ऐसी प्रयोग-अभिकल्पना की समस्या पर विचार करेंगे जिसका उद्देश्य इस कथन की सच्चाई जाँचना है।

§ १८१४२ प्रयोग-विधि

हमारा प्रयोग निम्नलिखित है। कुल आठ प्याले चाय बनायी जाय जिसमें से चार प्यालों में पहिले चाय और अन्य चार में पहिले दूध डाला जाय। इन प्यालों को महिला को एक यादृच्छिक क्रम से दिया जाय और वह चखकर यह बताने की चेष्टा करे कि कौन-सा पदार्थ पहिले डाला गया था—दूध या चाय। महिला को यह पहिले से बताया जाय कि प्रयोग में चार प्यालों में दूध पहिले और चार प्यालों में बाद में डाला गया है।

§ १८१४.३ अस्वीकृति प्रदेश और प्रतिदर्श परिमाण का निरचय

यह मालूम हो जाने के बाद स्वाभाविक ही है कि वह इन आठ प्यालों को चार चार के दो कुलों में इस प्रकार विभाजित करने की चेष्टा करेगी—एक में वह प्याले जिनमें दूध पहिले डाला गया है और दूसरे में वे जिनमें बाद में डाला गया है।

$$\text{आठ वस्तुओं में से चार वस्तुओं के कुल } \binom{8}{4} = \frac{8 \times 7 \times 6 \times 5}{4 \times 3 \times 2 \times 1} = 70$$

सचय बनाये जा सकते हैं। यदि महिला दोनों तरह के प्यालों में प्रभेद नहीं कर सकती तो उसके लिए अंदाज से इनको दो कुलों में ठीक-ठीक बाँटने की प्रायिकता $\frac{1}{70}$ है।

प्यालों की संख्या बढ़ाने से यह प्रायिकता और कम हो जाती है। इसके विपरीत यदि प्याला की संख्या को और छोटा कर दिया जाता तो यह प्रायिकता इतनी अधिक होती कि प्रयोग के फल को—यदि प्याला का प्रभेद ठीक भी हो गया हो—सयोग जनित माना जा सकता था। उदाहरण के लिए यदि केवल चार प्याले होते तो अंदाज से उन्हें दो सही सचचा में बाँटने की प्रायिकता $\frac{1}{\binom{4}{2}} = \frac{1 \times 2 \times 1}{4 \times 3} = \frac{1}{6}$ होती।

प्रयोगकर्ता को पहिले ही यह निश्चय कर लेना चाहिए कि वह क्या संख्या है जिससे कम प्रयोग के फल की प्रायिकता होने पर उसे विश्वास हो जायगा कि ऐसा केवल सयोग से नहीं हो सकता। इस प्रकार के प्रयोग से क्या लाभ जिसके किसी भी फल से उसे सतीय न हो। यदि वह यह सोचता है कि वे फल जिनकी प्रायिकता पाँच प्रतिशत अथवा उससे भी अधिक है किसी भी निष्कर्ष पर पहुँचने के लिए बेकार है तो उसके लिए आठ से कम प्यालों में प्रयोग करना निरर्थक है।

प्यालों की कोई भी संख्या सयोग के प्रभाव से हमें पूर्णतया नहीं बचा सकती। हम केवल इस सुविधाजनक नियम को मान लेते हैं कि यदि किसी घटना की प्रायिकता सत्तर में एक है तो वह सांख्यिकीय विचार से सार्थक है। आप यह तो समझ ही गये होंगे कि किसी एक प्रयोग से, चाहे उसका फल कितना ही सार्थक क्यों न हो, हमें पूर्ण विश्वास नहीं हो सकता। दस लाख में एक की प्रायिकता होने पर भी निश्चय ही वह घटना कभी न कभी घट ही सकती है। यह ही सकता है कि हमें आश्चर्य हो कि ऐसी असंभावी घटना हमारे ही प्रयोग में क्यों हुई।

यदि हम किसी प्राकृतिक घटना को प्रयोग द्वारा प्रमाणित करना चाहते हैं तो इसके-दुबके प्रयोग इसके लिए काफी नहीं हैं। इसके लिए भरपूर करने लायक एक विशेष प्रयोग-विधि की आवश्यकता है। मान लीजिए कि हमारे प्रयोग में महिला आठ में से छ प्यालों को ठीक-ठीक पहचान लेती है। यदि महिला में प्रभेद शक्ति नहीं हो तो इस घटना की प्रायिकता $\binom{4}{1} \binom{4}{1} - \binom{8}{2} = \frac{16}{70}$ है। यह स्पष्ट

है कि यदि इस घटना को सार्थक समझा जाता है, तो सही प्रभेद को तो सार्थक मानना ही पड़ेगा। इस प्रकार इस घटना अथवा इसमें अधिक सार्थक घटना के घटने की प्रायिकता $\frac{17}{70}$ है। यह बहुत अधिक है। इस कारण इस प्रयोग में केवल एक घटना

है जो साक्ष्यकीय दृष्टिकोण से सार्थक है और वह है महिला द्वारा प्यालो का शत प्रतिशत राही प्रभेद ।

§ १८१५ निराकरणीय परिकल्पना को सिद्ध नहीं किया जा सकता

इस प्रयोग में निराकरणीय परिकल्पना यह है कि महिला में प्रभेद शक्ति अनुपस्थित है । यह आपको याद ही होगा कि प्रयोग द्वारा निराकरणीय परिकल्पना को सिद्ध नहीं किया जा सकता—हाँ, उसका असिद्ध (disprove) होना समझ है । यह तर्क रखा जा सकता है कि यदि हमारा प्रयोग इस परिकल्पना को असिद्ध कर देता है कि महिला में प्रभेद शक्ति नहीं है, तो इसके द्वारा एक विपरीत कल्पना यह भी सिद्ध हो सकती है कि महिला में प्रभेद शक्ति विद्यमान है । परंतु यह विपरीत कल्पना एक निराकरणीय परिकल्पना का स्थान ग्रहण नहीं कर सकती, क्योंकि यह तो अनिश्चित ही रह जाता है कि विद्यमान प्रभेद शक्ति कितनी है । निराकरणीय परिकल्पना का पूर्णतः निश्चित (exact) होना आवश्यक है, क्योंकि इसके आधार पर ही प्रायिकता की गणना की जाती है ।

§ १८१६ भौतिक स्थितियों पर नियंत्रण की आवश्यकता

अब हमें यह देखना है कि किस दशा में यह कहा जा सकता है कि यदि महिला में प्रभेद शक्ति नहीं है तो प्रयोग के फल केवल मयोग पर निर्भर होंगे । मान लीजिए, उन सब प्यालो में जिनमें पहले दूध डाला जाता है, दो-दो चम्मच चीनी पड़ी हो, जब कि अन्य प्यालो में चीनी डाली ही नहीं गयी हो, तो दोनों प्रकार के प्यालो में प्रभेद करना बहुत ही आसान हो जायगा, क्योंकि यह स्वाद का भेद किरा भी मनुष्य द्वारा आसानी से पहचाना जा सकता है । इस प्रकार चार-चार प्यालो के में कुलक या तो सब ठीक या सब गलत श्रेणी में रखे जायेंगे और परिकल्पना की जाँच न्याययुक्त नहीं होगी । अतः प्रयोग में अन्य भौतिक स्थितियों पर नियंत्रण रखना भी आवश्यक है ।

§ १८१७ प्रयोग को अधिक सुग्राही (Sensitive) बनाने के कुछ तरीके

अब यदि महिला का कथन यह नहीं है कि वह हमेशा दो तरह के प्यालो में प्रभेद कर सकती है, बल्कि केवल यह है कि यद्यपि कभी कभी उससे मूल हो सकती है तथापि अधिकतर वह प्यालो को ठीक पहचान सकती है । इस दशा में उसको अपने कथन की सच्चाई का प्रमाण देने के लिए अधिक विस्तृत प्रयोग की आवश्यकता होगी ।

यदि प्रयोग में कुल बारह प्यालो का उपयोग किया जाय, जिनमें दोनो प्रकार के

□ -छ प्याले हों तो बिल्कुल ठीक प्रभेद करने की प्रायिकता $\frac{1}{\binom{12}{6}} = \frac{1}{924}$

है। 10 के ठीक और दो के गलत पहचानने जाने की प्रायिकता $\frac{\binom{6}{1}\binom{6}{1}}{\binom{12}{6}} = \frac{36}{924}$

है। क्योंकि $\frac{37}{924} < \frac{1}{20}$ इसलिए प्रयोग का यह फल भी सांख्यिकीय दृष्टिकोण से सार्थक माना जा सकता है। प्रयोगों के परिमाण को अधिकाधिक बढ़ाने से वह निराकरणीय परिकल्पना से प्राप्त तथा वास्तविक प्रायिकताओं के सूक्ष्मतर अंतर को पहचानने योग्य होता जाता है।

सूक्ष्मतर अंतर को पहचानने का एक और तरीका यह है कि छोटे प्रतिदर्श-परिमाण के प्रयोगों को ही कई बार दुहराया जाय। यदि आठ प्यालों के प्रयोग को ही आठ बार दुहराया जाय और इसमें से दो बार भी महिला ठीक प्रभेद कर पाये, तो इस घटना की और इससे भी अधिक सार्थक घटनाओं की प्रायिकता $1 - \left[\binom{8}{1} \times \frac{1}{70} \times \left(\frac{69}{70}\right)^7 + \binom{69}{70}^8 \right]$ है जो पाँच प्रतिशत से कम है। इस कारण इस फल को भी सार्थक माना जा सकता है।

प्रयोग को विस्तृत करने के अलावा उसे अधिक सुपाही बनाने के अन्य उपाय भी हैं। उदाहरण के लिए हर एक प्याले के लिए हम स्वतंत्र रूप से यह तय कर सकते थे कि उसमें दूध पहले डाला जाय या चाय। इसमें यह नियंत्रण उठा लिया गया है कि चार प्यालों में चाय पहले होगी और चार में दूध। हर एक प्याले को महिला के पान भेजने से पहले सिक्का उछालकर दूध या चाय के संबंध में निश्चय किया जा सकता है। यदि महिला में पभेद शक्ति नहीं है तो इस प्रकार भेजे हुए प्यालों को ठीक-ठीक पहचानने की प्रायिकता $\left(\frac{1}{2}\right)^8 = \frac{1}{256}$ है। सात प्यालों को ठीक और एक

को गलत बताने की प्रायिकता $\frac{8}{256} = \frac{1}{32}$ है जो पाँच प्रतिशत से कम है। इसलिए यह घटना भी सांख्यिकीय दृष्टिकोण से सार्थक है। इस प्रकार प्रयोग विधि को बदल देना कई बार लाभदायक होता है, परंतु इस विशेष प्रयोग में इस नूतन विधि का उपयोग कई बार गड़बड़ी पैदा कर सकता है। यह संभव है कि इस विधि के फलस्वरूप आठो प्याले एक ही प्रकार तैयार किये जायें। इस प्रकार के प्रयोग से जिस व्यक्ति पर

यह प्रयोग किया जा रहा हो उसका घबरा उठना स्वामाविक है। इसके अलावा यह हो सकता है कि यदि वह दोनो प्रकार की चाय चखे तो अतर को पहचान सकता है। परन्तु यदि सब प्यालो में एक ही प्रकार चाय बनायी जाय तो उसके पास इस अतर को पहचानने का कोई तरीका ही नहीं रह जाता।

ऊपर के प्रयोग को ध्यास्या से आप प्रयोग-परिमाण, यादृच्छिकीकरण तथा प्रयोग को नियंत्रण में रखने की आवश्यकता तथा महत्त्व समझ गये होंगे। हमें कई इसरो भी अधिक जटिल प्रयोगों का विश्लेषण करना होता है, जिनमें प्रायिकता इतनी सरलता से परिकल्पित नहीं हो सकती। इस काम के लिए कुछ अन्य सिद्धांतों की आवश्यकता होती है जिनको हम अगले कुछ अध्यायों में समझाने वा प्रयत्न करेंगे।

अध्याय १९

प्रसरण-विश्लेषण

(Analysis of Variance)

§ १९ १ एक प्रयोग

मान लीजिए कि एक कारखाने में खर के टुकड़े बनते हैं। किसी विशेष कार्य के लिए उनकी लंबाई एक निश्चित मान के लगभग होनी चाहिए। इन टुकड़ों की औसत लंबाई नापने के लिए एक प्रेक्षक रखा गया है। यह स्पष्ट है कि प्रेक्षक यदि हर एक टुकड़े को नापे तो बहुत अधिक समय लगेगा। इसलिए वह कारखाने में बने हुए खर के टुकड़ों के एक प्रतिदर्श को लेकर उसी की लंबाई नापेगा। इसके अलावा एक ही टुकड़े की लंबाई भी यदि बार-बार नापी जाय तो फल हमेशा एक-सा नहीं होगा। कुछ तो इस कारण कि मापनी (scale) के दो विभाजनों के बीच में होने पर प्रेक्षक को अनुमान लगाना पड़ता है। इसके अलावा खर की लंबाई को नापने के लिए उसे खींचकर रखना पड़ता है। इस खिंचाव से भी लंबाई में अंतर पड़ सकता है और यदि प्रयोग बार-बार किया जाय तो खिंचाव हर बार बिल्कुल एक-सा नहीं होगा।

इस प्रकार यदि एक प्रतिदर्श से टुकड़ों की औसत लंबाई का अनुमान लगाया जाता है तो उसमें दो प्रकार की त्रुटियाँ का प्रभाव पड़ेगा। एक तो भिन्न भिन्न टुकड़ों की लंबाई में अंतर के कारण और दूसरे एक ही टुकड़े की लंबाई के नापने में प्रेक्षक त्रुटि (observational error) के कारण। इसी प्रकार लगभग सभी प्रयोगों का फल अनेक उपादानों पर निर्भर करता है। कई बार प्रयोग का उद्देश्य यह जानना होता है कि किसी विशेष उपादान का कोई प्रभाव है या नहीं।

§ १९ २ प्रसरणों का सयोज्यता गुण (Additive property of variances)

ऊपर के प्रयोग में टुकड़ों की प्रेषित लंबाइयाँ यादृच्छिक खर हैं। मान लीजिए कि कुल k टुकड़ों का प्रतिदर्श चुना गया है। इनमें से 1 -वें टुकड़े की लंबाई को हम 1 , से सूचित करेंगे। यदि समाप्ति के कुल टुकड़ों की औसत लंबाई $\bar{1}$ हो तो एक त्रुटि

तो समष्टि में से केवल k टुकड़ों के चुने जाने के कारण होगी, जो प्रतिदर्श-परिमाण और l के प्रसरण पर निर्भर करेगी। इस त्रुटि को प्रतिदर्श-त्रुटि (sampling error) कहते हैं। यह प्रसरण $E(l_i - l)^2$ है जिसको हम σ_1^2 से सूचित करेंगे।

मान लीजिए, प्रतिदर्श के i -वें टुकड़े को n_i बार नापा जाता है और j -वीं बार के नापने के फल को l_{ij} से सूचित करते हैं। l_{ij} भी एक यादृच्छिक चर है जिसके प्रसरण $E[(l_{ij} - l_i)^2 | l_i]$ को हम σ_0^2 से सूचित करेंगे। हम यह मान लेते हैं कि यह प्रसरण, जो प्रेक्षण त्रुटि का माप है, हर एक टुकड़े के लिए बराबर है। यदि हम बिना प्रतिबन्ध के l_{ij} के प्रसरण को σ^2 से सूचित करें तो

$$\begin{aligned} \sigma^2 &= E[l_{ij} - l]^2 \\ &= E[(l_{ij} - l_i) + (l_i - l)]^2 \\ &= E(l_{ij} - l_i)^2 + E(l_i - l)^2 \\ &= \sigma_0^2 + \sigma_1^2 \end{aligned} \quad (191)$$

इस प्रकार त्रुटियों के उद्गम यदि स्वतंत्र रूप से प्रभाव डालते हैं तो जो कुल प्रसरण इन दोनों उद्गमों के संयुक्त प्रभाव से होता है, वह अलग-अलग प्रभावों के प्रसरणों का योग होता है।

इस गुण को प्रसरणों का संयोग्यता गुण कहते हैं।

§ १९३ औसत लंबाई का प्राक्कलन

अब हम देखें कि कुल टुकड़ों की औसत लंबाई का अनुमान कैसे लगाया जा सकता है। हमें यह पता है कि l_{ij} का प्रत्याशित मान l है। यह इस कारण कि

$$\begin{aligned} E(l_{ij}) &= E[E(l_{ij} | l_i)] \\ &= E[l_i] \\ &= l \end{aligned}$$

इस प्रकार यादृच्छिकीकरण द्वारा चुने हुए हर एक टुकड़े पर लिया हुआ प्रत्येक प्रेक्षण l_{ij} समष्टि में औसत लंबाई का अनभिन्न प्राक्कलक है। इस कारण यदि

$$\sum_{i=1}^k \sum_{j=1}^{n_i} k_{ij} = 1 \text{ हो जहाँ प्रत्येक } k_{ij} \text{ एक अचर मर्यादा है तो } \sum_{i=1}^k \sum_{j=1}^{n_i} k_{ij} l_{ij} \text{ भी } l$$

का एक अनभिन्न प्राक्कलक है क्योंकि

$$\begin{aligned}
 E\left[\sum_{i=1}^k \sum_{j=1}^{n_i} l_{ij}\right] &= \sum_{i=1}^k \sum_{j=1}^{n_i} l_{ij} E(l_{ij}) \quad (\text{देखिए § ४१०}) \\
 &= l \sum_{i=1}^k \sum_{j=1}^{n_i} k_{ij} \\
 &= l
 \end{aligned}$$

क्योंकि सब l_{ij} प्रेक्षणों का प्रसरण बराबर है, इस कारण इन चरों का वह एक-आती फलन जिसका प्रसरण निम्नतम हो ऐसा होना चाहिए कि उसमें सब l_{ij} वाले पदों के गुणक बराबर हों। इसलिए इन प्रेक्षणों पर आधारित सर्वोत्तम प्राक्कलक होगा

$$T = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} l_{ij}}{n}$$

$$\text{जहाँ } n = \sum_{i=1}^k n_i$$

§ १९४ औसत लंबाई के प्राक्कलक का प्रसरण

इस प्राक्कलक का प्रसरण क्या होगा? इसके लिए हम निम्नलिखित सिद्धांत का उपयोग करते हैं। यदि एक ही टुकड़े—मान लीजिए १—वें टुकड़े—को ही n_1 बार नापा जाय और इन प्रेक्षणों के माध्य को कुल टुकड़ों की लंबाई के माध्य का अनुमान समझा जाय तो इसमें प्रेक्षण त्रुटि तो कम होकर $\frac{\sigma_0^2}{n_1}$ रह जायगी, परंतु प्रतिदर्शी त्रुटि में कुछ बन्नी नहीं आवेगी। इस प्रकार इस अनुमान का प्रसरण $\sigma_1^2 + \frac{\sigma_0^2}{n_1}$ होगा।

यदि इस अनुमान को \bar{l}_1 से सूचित किया जाय तो

$$V(\bar{l}_1) = \sigma_1^2 + \frac{\sigma_0^2}{n_1} \quad (19.2)$$

$$\text{परंतु } T = \frac{1}{n} \sum_{i=1}^k n_i \bar{l}_i \quad \text{और } \bar{l}_1, \bar{l}_2, \dots, \bar{l}_k$$

सब स्वतंत्र चर हैं। इसलिए

$$\begin{aligned}
 V(\bar{l}) &= \frac{1}{n^2} \sum_{i=1}^k n_i^2 \left[\sigma_1^2 + \frac{\sigma_0^2}{n_i} \right] \\
 &= \sum_{i=1}^k \frac{n_i^3}{n^2} \sigma_1^2 + \frac{1}{n^2} \sum_{i=1}^k n_i \sigma_0^2 \\
 &= \frac{\sigma_1^2}{n^2} \sum_{i=1}^k n_i^2 + \frac{\sigma_0^2}{n} \qquad \dots\dots(19.3)
 \end{aligned}$$

यदि सब टुकड़ों पर प्रेक्षणों की संख्या बराबर हो और हम इस संख्या को m से सूचित करें तो

$$\begin{aligned}
 n_i &= m, \quad n = m k \\
 \therefore V(\bar{l}) &= \frac{1}{k} \sigma_1^2 + \frac{1}{mk} \sigma_0^2 \qquad \dots\dots(19.4)
 \end{aligned}$$

§ १९.५ प्रसरण का प्राक्कलन

जब हम किसी प्राचल का अनुमान लगाते हैं तो यह भी आवश्यक है कि हमें इस अनुमान की त्रुटि का भी कुछ अंदाजा हो। यानी हमें $V(\bar{l})$ के प्राक्कलन की भी आवश्यकता है। हम कोशिश करेंगे कि हमें σ_1^2 तथा σ_0^2 के अलग अलग प्राक्कलन प्राप्त हो जायें।

§ १९.५१ σ_0^2 का प्राक्कलन

आइए, पहिले हम यह देखें कि σ_0^2 का क्या प्राक्कलक हो सकता है। क्योंकि इसमें हम प्रेक्षणों की त्रुटि का पता चलाना चाहते हैं, यह प्राक्कलक एक ही टुकड़े की विभिन्न प्रेक्षित लब्धाद्यों के अंतर से संबंधित होना चाहिए। मान लीजिए कि हम i -वें टुकड़े पर किये हुए प्रेक्षणों को ही ध्यान में रखते हैं। इन प्रेक्षणों की त्रुटियों का

वर्ग-योग $\sum_{j=1}^{n_i} (l_j - \bar{l}_i)^2$ है।

$$E \left[\sum_{j=1}^{n_i} (l_j - \bar{l}_i)^2 \right] = E \left[\sum_{j=1}^{n_i} \{ (l_j - l) - (\bar{l}_i - l) \}^2 \right]$$

$$\begin{aligned}
 &= \sum_{j=1}^{n_1} E(l_{1j} - l_1)^2 - n_1 E(\bar{l}_1 - l_1)^2 \\
 &= n_1 \sigma_0^2 - n_1 \frac{\sigma_0^2}{n_1} \\
 &= \sigma_0^2 (n_1 - 1) \qquad (19.5)
 \end{aligned}$$

इस प्रकार σ_0^2 का एक अनभिन्नत प्राक्कलन $\frac{\sum_{j=1}^{n_1} (l_{1j} - \bar{l}_1)^2}{n_1 - 1}$ है। इस प्रकार

विभिन्न टुकड़ा से σ_0^2 का प्राक्कलन किया जा सकता है। इन विभिन्न प्राक्कलकों का भारित माध्य (weighted mean) भी σ_0^2 का अनभिन्नत प्राक्कलक होगा।

उदाहरण के लिए $M_0 = \frac{S_0}{n k} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (l_{ij} - \bar{l}_i)^2}{n - k}$ इसी प्रकार का एक भारित माध्य है जिसमें i -वें प्राक्कलक का भार $(n_i - 1)$ है।

परन्तु $\sum_{i=1}^k (n_i - 1) = (n - k)$ है।

§ १९.५.२ σ_1^2 का प्राक्कलन

इस प्रकार हम प्रेक्षण त्रुटि का अनुमान लगा सकते हैं। आइए अब हम देखें कि प्रतिदर्शी त्रुटि σ_1^2 का अनुमान किस प्रकार लगाया जाय। क्योंकि यह त्रुटि टुकड़ों की वास्तविक लंबाइयों का प्रसरण है, इसलिए यह स्वाभाविक है कि हम इसके लिए टुकड़ा पर किये प्रेक्षणों के माध्यों के अंतर की परीक्षा करें। उदाहरण के लिए

$$\begin{aligned}
 S_1 &= \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{l}_i - \bar{l})^2 \\
 &= \sum_{i=1}^k n_i \bar{l}_i^2 - n \bar{l}^2 \qquad (19.6)
 \end{aligned}$$

$$E(S_1) = \sum_{i=1}^k n_i \left[\sigma_1^2 + \frac{\sigma_0^2}{n_i} + I^2 \right] - n \left[\frac{\sigma_1^2}{n} \sum_{i=1}^k n_i^2 + \frac{\sigma_0^2}{n} + I^2 \right]$$

$$= \sigma_1^2 \left[n - \frac{\sum_{i=1}^k n_i^2}{n} \right] + (k-1) \sigma_0^2 \quad \dots (197)$$

क्याकि $E(\bar{l}_i) = V(\bar{l}_i) + I^2, E(\bar{l}^2) = V(\bar{l}) + I^2$ तथा $\sum_{i=1}^k n_i = n$, इस प्रकार σ_1^2

का प्राक्कलक $S'_1 = \frac{S_2 - (k-1)M_0}{n - \frac{\sum_{i=1}^k n_i^2}{n}}$ होगा। यदि सब n_i बराबर हो और

एतका मान m हो तो

$$S'_1 = \frac{S_2 - (k-1)M_0}{n-m} \quad \dots (198)$$

$$\text{तथा } S_1 = m \sum_{i=1}^k (\bar{l}_i - \bar{l})^2 \quad \dots (199)$$

§ १९६ प्रसरण विश्लेषण (Analysis of variance)

इस प्रसरणो के प्राक्कलनो के कलन के लिए यह ध्यान देने योग्य बात है कि

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (l_{ij} - \bar{l})^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (l_{ij} - \bar{l}_i)^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{l}_i - \bar{l})^2$$

$$= S_0 + S_1 \quad \dots (1910)$$

इस प्रकार साधारण माध्य \bar{l} से प्रेक्षणो के वर्ग विचलनो (squared deviations) का योग दो भागो के योग के रूप में रखा जा सकता है—

(१) समूह माध्य (group mean) से उस समूह के समस्त चरो के वर्गित विचलनो का योग जिसको समूहाम्यन्तरिक वर्ग-योग (within group sum of squares) कहा जा सकता है।

(२) साधारण माध्य से समूह-माध्यों के वर्गित विचलनों का योग, जिसको अंतः-सामूहिक वर्ग-योग (between group sum of squares) की संज्ञा दी जा सकती है, अर्थात्

$$\text{सम्पूर्ण वर्ग-योग} = \text{अंतर-सामूहिक वर्ग-योग} + \text{समूहाम्यन्तरिक वर्ग-योग} \quad \dots\dots(19\ 11)$$

इस प्रकार सम्पूर्ण वर्गित विचलन योग को कुछ भागों में विभाजित करने को प्रसरण विश्लेषण कहते हैं।

§ १९.७ प्रसरण विश्लेषण का परिकल्पना की जाँच में उपयोग

दो प्रकार की समस्याएँ हैं जिनमें प्रसरण विश्लेषण का उपयोग होता है। एक में तो प्रेक्षणों को कुल सभ्य प्रेक्षणों के एक काल्पनिक जगत् का प्रतिदर्श मान लिया जाता है। विश्लेषण का उद्देश्य इस जगत् के प्रसरण का प्राक्कलन करना होता है। यह कैसे किया जा सकता है यह हम ऊपर के उदाहरण में देख ही चुके हैं। जिन रबर के टुकड़ों को नापा जाता है वह कुल रबर के टुकड़ों के जगत् का एक यादृच्छिकीकृत प्रतिदर्श है। एक ही टुकड़े के जितने नाप लिये जाते हैं उनके कुलक को उस टुकड़े के सब सभ्य नापों के एक काल्पनिक जगत् का प्रतिदर्श माना जाता है। इन दो जगत्ओं के प्रसरण क्रमशः σ_1^2 और σ_0^2 हैं और उद्देश्य इन दोनों प्रसरणों का अनुमान लगाना है। दूसरे प्रकार की समस्या होती है माध्यों की तुलना। यदि दो समष्टियाँ हो और निराकरणिय परिकल्पना यह हो कि इन दोनों के माध्य समान हैं तो इसकी जाँच किस प्रकार की जायेगी यह हम पहिले ही देख चुके हैं। यदि हमें दो नहीं बल्कि अनेक समष्टियों के माध्यों की तुलना करनी हो अथवा इस परिकल्पना की जाँच करनी हो कि इन सब समष्टियों के माध्य बराबर हैं तो हमें प्रसरण विश्लेषण की सहायता लेनी पड़ती है।

मान लीजिए कि ऊपर के उदाहरण में हमारी निराकरणिय परिकल्पना यह है कि प्रतिदर्शों के प्रत्येक टुकड़े की वास्तविक लंबाई बराबर है। यदि ऐसा हो तो $\sigma_1^2 = 0$ और

$$E(S_1) = (k-1) \sigma_0^2 \quad \dots\dots (19\ 12)$$

[दिए गए समीकरण (19 7)]

$$\text{इस प्रकार परिकल्पना के अंतर्गत } M_0 = \frac{S_0}{n-k} \text{ तथा } M_1 = \frac{S_1}{k-1}$$

दोनों ही σ_0^2 के अनमिनत प्राक्कलक हैं। परंतु यदि परिकल्पना सत्य न हो तो M_1 का प्रत्याशित मान σ_0^2 से अधिक होता है। इस कारण यदि यह मान लें कि

$$F = \frac{M_1}{M_0} = \frac{S_1/(k-1)}{S_0/(n-k)}$$

$$= \frac{(\text{अंतर-सामूहिक वर्गयोग})/(k-1)}{(\text{समूहान्तरिक वर्गयोग})/(n-k)} \quad \dots \dots (19.13)$$

तो F ऐसा चर है जिसका मान परिकल्पना की सत्यता पर रोशनी डाल सकता है । यदि यह बहुत अधिक हो तो परिकल्पना पर शक होना स्वाभाविक ही है ।

§ १९.८ प्रसरण-विश्लेषण सारणी (Analysis of variance table)

अंतर सामूहिक, समूहान्तरिक और सम्पूर्ण वर्ग-योगों और उनकी स्वातंत्र्य सख्याओं को एक सारणी के रूप में रखा जा सकता है । इस सारणी को प्रसरण विश्लेषण सारणी कहते हैं । ऊपर के प्रयोग के लिए हमें जो सारणी प्राप्त होती है वह नीचे दी हुई है ।

सारणी संख्या 19.1

प्रसरण विश्लेषण सारणी

विचरण	वर्ग-योग	स्वातंत्र्य सख्या	वर्ग-माध्य	वर्ग-माध्य का प्रत्याशित मान
(1)	(2)	(3)	(4)	(5)
अंतर सामूहिक	$\sum_{j=1}^k n_j (\bar{l}_j - \bar{l})^2$ $= S_1$	$k-1$	$\frac{S_1}{k-1} = M_1$	$\sigma_0^2 + \sigma_1^2 \left[n - \frac{\sum_{j=1}^k n_j^2}{n} \right]$
समूहान्तरिक	$\sum_{j=1}^k \sum_{l=1}^{n_j} (l_{jl} - \bar{l}_j)^2$ $= S_0$	$n-k$	$\frac{S_0}{n-k} = M_0$	σ_1^2
सम्पूर्ण	$\sum_{j=1}^k \sum_{l=1}^{n_j} (l_{jl} - \bar{l})^2$ $= S$	$n-1$	$\frac{S}{n-1} = \frac{S_1 + S_0}{n-1} = M$	$\sigma_0^2 + \frac{k-1}{n-1} \left[n - \frac{\sum_{j=1}^k n_j^2}{n} \right] \sigma^2$

इस सारणी द्वारा यह सरलता से देखा जा सकता है कि सम्पूर्ण वर्ग-योग अंतर-सामूहिक और समूहान्तरिक वर्ग-योगों का योग है । इसी प्रकार कुल स्वातंत्र्य सख्या भी अंतर-सामूहिक और समूहान्तरिक स्वातंत्र्य-सख्याओं का योग है । वर्ग-

योगी का यह सयोज्यता-गुण प्रसरण विस्लेषण में बहुत महत्त्वपूर्ण है। यदि हम अतर सामूहिक वर्ग-योग तथा सम्पूर्ण वर्ग-योग का कलन कर लें तो समूहाम्यन्तरिक वर्ग-योग पहले को दूसरे में से घटा कर मालूम विया जा सकता है। प्रसरण विस्लेषण सारणी का उद्देश्य केवल इस प्रकार से समूहाम्यन्तरिक वर्ग-योग का कलन ही नहीं बल्कि अत में वर्ग-माध्यों के अनुपात $F = \frac{M_1}{M_0}$ का परिवर्तन है। यही वह धर है जिसके मान के आधार पर हमें सब समूहों के माध्य के बराबर होने की परिवर्तना की जाँच करनी है।

§ १९९ कुछ कल्पनाएँ जिनके आधार पर निराकरणिय परिवर्तना की जाँच की जाती हैं

(1) मान लीजिए कि i -वें समूह पर किया हुआ j -वाँ प्रेक्षण l_{ij} एक प्रसामान्य चर है जिसका माध्य l_i और प्रसरण σ_{oi}^2 है। इस दशा में हम l_{ij} को निम्नलिखित रूप में रख सकते हैं।

$$l_{ij} = l_i + \epsilon_{ij}$$

जहाँ ϵ_{ij} एक प्रसामान्य चर है जिसका माध्य 0 और प्रसरण σ_{oi}^2 है।

(2) यदि ये ϵ_{ij} एक दूसरे से स्वतंत्र हों तो

$$\frac{1}{\sigma_{oi}^2} \sum_{j=1}^{n_i} (l_{ij} - l_i)^2 = \frac{1}{\sigma_{oi}^2} \sum_{j=1}^{n_i} (\epsilon_{ij} - \bar{\epsilon}_i)^2$$

ऐसा χ^2 -चर होगा जिसकी स्वातन्त्र्य-सख्या $(n_i - 1)$ है। (देखिए § १९१)

$$\text{इसी प्रकार } \frac{1}{\sigma_{oi}^2} \sum_{j=1}^{n_i} (l_{ij} - l_i)^2, \frac{1}{\sigma_{o2}^2} \sum_{j=1}^{n_2} (l_{2j} - l_2)^2,$$

आदि सब यादृच्छिक चरों के वटन भी χ^2 वटन हैं जिनकी स्वातन्त्र्य सख्याएँ क्रमशः $(n_1 - 1), (n_2 - 1)$..इत्यादि हैं। इसके अलावा ये धर एक दूसरे से स्वतंत्र हैं।

इस कारण इन सबका योग $\sum_{i=1}^k \frac{1}{\sigma_{oi}^2} \sum_{j=1}^{n_i} (l_{ij} - l_i)^2$ भी एक χ^2 -चर है जिसकी

स्वानम्य सख्या $\sum_{i=1}^k (n_i - 1) = (n - k)$ है। (देखाएँ § ९.४)

(३) अब यहाँ एक और कल्पना करते हैं। वह यह कि हर एक टुकड़े के लिए प्रेक्षण-प्रसरण बराबर है। यानी

$$\sigma_{01}^2 = \sigma_{02}^2 = \dots = \sigma_{0n}^2 = \sigma_0^2$$

$$\text{इसलिए } \frac{S_0}{\sigma_0^2} = \frac{1}{\sigma_0^2} \sum_{i=1}^k \sum_{j=1}^{n_i} (l_{ij} - \bar{l}_i)^2 \text{ भी एक } \chi_{2-k}^2 \text{ चर है।}$$

इसके अलावा

$$(\bar{l}_i - \bar{l}) = (l_i - \bar{l}) + (\bar{e}_i - \bar{e})$$

यहाँ \bar{e}_i एक $N\left(0, \frac{\sigma_0}{\sqrt{n_i}}\right)$ चर है। इस कारण

$\bar{e}_1, \bar{e}_2, \dots, \bar{e}_k$ का भारित प्रसरण

$$\frac{1}{\sigma_0^2} \sum_{i=1}^k n_i (\bar{e}_i - \bar{e})^2 = \text{एक } \chi_{k-1}^2 \text{ चर है। परन्तु}$$

$$\sum_{i=1}^k n_i (\bar{e}_i - \bar{e})^2 = \sum_{i=1}^k n_i [(\bar{l}_i - \bar{l}) - (l_i - \bar{l})]^2$$

§ १९१० F-परीक्षण

यदि हमारी निराकरणिय परिवर्तना यह हो कि

$$l_1 = l_2 = \dots = l_k = l \text{ तो}$$

$$\sum_{i=1}^k n_i (\bar{e}_i - \bar{e})^2 = \sum_{i=1}^k n_i (\bar{l}_i - \bar{l})^2 = S_1$$

इसलिए इस परिवर्तना के अतर्गत अंतर-सामूहिक प्रसरण $\frac{S_1}{n_0^2}$ एवं

χ_{k-1}^2 चर है और क्योंकि यह S_0 से स्वतंत्र है इस कारण इस परिवर्तना के अतर्गत

$$F = \frac{S_1/k-1}{S_0/n-k} \text{ एक } F_{k-1, n-k} \text{ चर है। (देखाएँ § १११)}$$

यदि इसका प्रेक्षित मान सारणी में दिये हुए $F_{n-1, n-2}$ के पाँच प्रतिशत बिंदु अथवा किसी निश्चित बिंदु से अधिक हो तो हम इस परिकल्पना को मूलतः समझते हैं।

ऊपर हमने देखा कि कुछ परिकल्पनाओं के अंतर्गत दो वर्ग-माध्यों का अनुपात एक F -चर होता है और इस कारण हम उन परिकल्पनाओं की जाँच प्रयोग द्वारा कर सकते हैं। ऊपर यह सिद्ध करने के लिए कि इस अनुपात का वटन F -वटन है हमने प्रसापान्यता आदि कुछ अन्य कल्पनाओं को भी अपनी मुख्य परिकल्पना के साथ मिला दिया था। सांख्यिकी में गणना करके यह सिद्ध कर दिया है कि इन अन्य कल्पनाओं की अनुपस्थिति में यद्यपि चर का वटन F -वटन नहीं होगा, परंतु उसके वास्तविक वटन की 95 प्रतिशत विश्वास्य सीमाएँ F -वटन की 95 प्रतिशत विश्वास्य सीमाओं से इतने कम अंतर पर होंगी कि हम F -वटन का ही प्रयोग परिकल्पना को जाँचने के लिए यदि करें तो कोई विधेय त्रुटि नहीं होगी।

इस उदाहरण में हमने देखा कि दो प्रकार की त्रुटियों में से एक प्रकार की त्रुटि की अनुपस्थिति की परिकल्पना को कैसे जाँचा जाता है। अन्य कई ऐसी परिस्थितियाँ हो सकती हैं जिनमें कई प्रकार की त्रुटियाँ प्रेक्षण पर प्रभाव डालती हैं। इस स्थिति में हम बारी-बारी से हर एक की अनुपस्थिति की परिकल्पना की जाँच करना चाहेंगे। इसके लिए यह आवश्यक नहीं है कि विचरण के प्रत्येक उद्गम की प्रभावशीलता की जाँच के लिए एक नया प्रयोग किया जाय। प्रयोग की अभिकल्पना इस प्रकार की जा सकती है कि एक ही प्रयोग में सब परिकल्पनाओं की जाँच हो सके। आगे के अध्यायों में इस प्रकार की कुछ अभिकल्पनाओं को उदाहरण सहित समझाने की चेष्टा की गयी है।

अध्याय २०

यादृच्छिकीकृत-ब्लॉक अभिकल्पना (Randomized Block Design)

§ २०१ ब्लॉक बनाने का उद्देश्य

मान लीजिए, गेहूँ की चार किस्में हैं और हम प्रयोग द्वारा यह जानना चाहते हैं कि इनमें से सर्वोत्तम कौन-सी है। यहाँ अच्छी किस्म से हमारा तात्पर्य उस किस्म से है जिसमें प्रति एकड़ अधिक गेहूँ उत्पन्न हो। यह कहा जा सकता है कि यह प्रयोग तो अत्यन्त सरल है। आप इन विभिन्न किस्मों को जोकर देख लीजिए कि किसमें गेहूँ अधिक होता है। परन्तु भाइए हम तनिक ध्यान इस बात पर दें कि इस प्रयोग में क्या क्या विकर्ण हो सकती हैं। सबसे बड़ी और पहली दिक्कत तो यह है कि गेहूँ की डगज केवल उसकी किस्म पर ही निर्भर नहीं करती बल्कि बहुत हद तक जमीन भी इसको प्रभावित करती है। यदि धरती उपजाऊ हो तो उसमें गाम्भीर्य किस्म का गेहूँ भी अधिक उपज दे सकता है। यदि इन प्रयोग में सयोग से अच्छी किस्म का गेहूँ बजर धरती में बो दिया गया और मामूली किस्म का गेहूँ उपजाऊ धरती में बोया जाय तो यह सभन है कि प्रयोग से निष्कर्ष उलटा ही निकले। इसलिए इस बात का ध्यान रखना पड़ेगा कि सब गेहूँ एक समान उपजाऊ धरती में बोये जायें। परन्तु खाद इत्यादि देकर तथा ऊपर से हल चलाकर और पानी देकर खेतों को एक समान करने की चाहे जितनी चेष्टा की जाय उनमें कुछ न कुछ अंतर रह ही जायगा।

यदि आप यह सोचते हों कि एक ही खेत में बारी बारी में किस्मों को बोने से यह समस्या हल हो जायगी तो यह भी आपका भ्रम है। एक तो यह दिक्कत है कि परती का उपजाऊपन समय के साथ बदलता है और किसी हद तक इस बात पर निर्भर करता है कि पिछले वर्ष इसमें कौन-सी फसल बोयी गयी थी। इसके अलावा जलवायु का खेती पर जो महत्वपूर्ण प्रभाव पड़ता है उसे तो आप जानते ही हैं। इसके ही कारण एक ही खेत में एक ही प्रकार के गेहूँ की उपज भी भिन्न-भिन्न वर्षों में भिन्न-भिन्न होती है। इसलिए यदि हमें गेहूँ को किस्मों की तुलना करनी है तो यह आवश्यक है कि प्रयोग-काल अलग-अलग न हो।

इस प्रकार हम इस निष्कर्ष पर पहुँचते हैं कि एक ही समय में और जहाँ तक हो सके एक समान उपजाऊ धरती पर ही इन सब किस्मों को बोया जाय । यदि एक ही खेत के छोटे-छोटे विभाजन करके उसमें उनको बोया जाय तो यह आशा की जा सकती है कि इन विभाजनों के उपजाऊपन में विशेष अंतर नहीं होगा । फिर भी कुछ अंतर इनमें अवश्य होगा और इसका ध्यान हमें तुलना करते समय रखना पड़ेगा । यदि प्रेक्षित उपजों का अंतर साधारण हो तो कदाचित् यह इन विभाजनों के उपजाऊपन के अंतर के कारण ही हो और इस परिस्थिति में हमारे लिए यह कहना संभव नहीं है कि कौन-सी किस्म सर्वश्रेष्ठ है अथवा किस्मों की उपज में कुछ अंतर है भी अथवा नहीं ।

§ २० २ यादृच्छिकीकरण और पुनः प्रयोग (*Randomization and replication*)

किमी विशेष किस्म की कोई तरफदारी हम अपनी ओर से नहीं करना चाहते । इसलिए किस विभाजन में कौन-सी किस्म का गेहूँ बोया जाय, यह निश्चय यादृच्छिकीकरण द्वारा किया जाता है । फिर भी संयोग के प्रभाव को कम करने के लिए यह आवश्यक है कि एक ही किस्म का गेहूँ एक से अधिक विभाजन में बोया जाय । इस प्रकार यदि संयोग से एक विभाजन उसे अच्छा मिल जाता है तो एक साधारण भी मिले । सभी विभाजन अच्छे या सभी साधारण हो इसकी प्रायिकता को घटा कर हम लगभग धूम्य के बराबर कर देना चाहते हैं । इसके लिए जो तरीक़ब साधारणतया काम में लामी जाती है वह निम्नलिखित है ।

एक साधारण लंबाई चौड़ाई के भूमि खंड को, जिसे आगे हम ब्लॉक कहेंगे, चार भागों में विभाजित किया जाता है । इन भागों को हम प्लॉट कहेंगे । इन चारों भागों में एक-एक किस्म का गेहूँ बो दिया जाता है । कौन-से प्लॉट में कौन सा गेहूँ बोया जायगा, यह यादृच्छिकीकरण द्वारा तय किया जाता है । इन प्लॉटों के एक छोटे भूखण्ड के भाग होने के कारण समझा जा सकता है कि इनके स्वाभाविक उपजाऊपन में अधिक अंतर होगा । इस प्रकार के भिन्न-भिन्न कई ब्लॉकों में प्रयोग किया जाता है जिनमें से हर एक में गेहूँ की चार किस्मों के लिए प्लॉटों का वितरण यादृच्छिकीकरण द्वारा किया जाता है ।

§ २० ३ यादृच्छिकीकृत ब्लॉक अभिकल्पना और पूर्णतः यादृच्छिकीकृत अभिकल्पना में अन्तर

इस प्रकार यदि कुल r ब्लॉकों पर प्रयोग किया जाय तो प्रत्येक प्रकार के गेहूँ के लिए r प्लॉट मिलते हैं । परंतु यह $4r$ प्लॉटों में से r प्लॉटों के यादृच्छिकीकरण

द्वारा चुने जाने से भिन्न है। इस प्रकार की पूर्णतः यादृच्छिकीकृत अभिकल्पना (completely randomized design) में इस घटना की प्रायिकता बहुत कम है कि हर एक ब्लॉक में हर एक किस्म का गेहूँ एक-एक प्लॉट में बोया जाय। किसी ब्लॉक में किसी विशेष किस्म का गेहूँ दो या अधिक बार बोया जाता और किसी अन्य ब्लॉक में किसी अन्य किस्म का। यह संभव है कि ब्लॉकों के उपजाऊन में काफी अंतर हो। इस दशा में यदि इन चार किस्मों के गेहूँ की औसत पैदावारी की तुलना प्रयोग में की जाय तो उसमें ब्लॉकों के उपजाऊन का अंतर इतनी अधिक भ्रुति उत्पन्न करेगा कि यदि इन किरगों में अंतर गामूली हो तो इस प्रयोग द्वारा हम इसे नहीं जान सकेंगे। परन्तु हर एक ब्लॉक में प्रत्येक किस्म के गेहूँ को एक एक प्लॉट में बोने से यदि ब्लॉकों के बीच में कुछ अंतर हो भी तो उसका प्रभाव जाता रहता है। इन प्रकार की प्रयोग अभिकल्पना को यादृच्छिकीकृत-ब्लॉक अभिकल्पना (randomized block design) कहते हैं।

नीचे इसी प्रकार के प्रयोग का एक नक्शा दिया हुआ है। चार किस्म के गेहूँओं को क्रमशः A, B, C और D की संज्ञा दी गयी है। ब्लॉकों को नम्बर I, II इत्यादि दिये गये हैं। इस प्रयोग में ब्लॉकों की कुल संख्या छ है।

<p>I</p> <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr> <td style="padding: 10px;">A</td> <td style="padding: 10px;">D</td> </tr> <tr> <td style="padding: 10px;">C</td> <td style="padding: 10px;">B</td> </tr> </table>	A	D	C	B	<p>II</p> <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr> <td style="padding: 10px;">B</td> <td style="padding: 10px;">C</td> </tr> <tr> <td style="padding: 10px;">A</td> <td style="padding: 10px;">D</td> </tr> </table>	B	C	A	D	<p>III</p> <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr> <td style="padding: 10px;">C</td> <td style="padding: 10px;">A</td> </tr> <tr> <td style="padding: 10px;">D</td> <td style="padding: 10px;">B</td> </tr> </table>	C	A	D	B
A	D													
C	B													
B	C													
A	D													
C	A													
D	B													
<p>IV</p> <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr> <td style="padding: 10px;">C</td> <td style="padding: 10px;">D</td> </tr> <tr> <td style="padding: 10px;">A</td> <td style="padding: 10px;">B</td> </tr> </table>	C	D	A	B	<p>V</p> <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr> <td style="padding: 10px;">C</td> <td style="padding: 10px;">D</td> </tr> <tr> <td style="padding: 10px;">A</td> <td style="padding: 10px;">B</td> </tr> </table>	C	D	A	B	<p>VI</p> <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr> <td style="padding: 10px;">B</td> <td style="padding: 10px;">D</td> </tr> <tr> <td style="padding: 10px;">A</td> <td style="padding: 10px;">C</td> </tr> </table>	B	D	A	C
C	D													
A	B													
C	D													
A	B													
B	D													
A	C													

§ २०.४ वे उपादान जिन पर पैदावार निर्भर करती है

किसी भी प्लॉट में गेहूँ की पैदावार तीन चीजों पर निर्भर करती है।

(१) गेहूँ की किस्म,

(२) ब्लॉक की भूमि का उपजाऊपन,

(३) ब्लॉक के अंदर का वह प्लॉट जिस पर यह किस्म बोयी गयी है। यह

अंतिम चुनाव यादृच्छिकीकृत होने के कारण हम इस प्लॉट-प्रभाव का वटन मालूम कर सकते हैं। इसलिए किस्मों के अंतर की जांच करने के लिए यह आवश्यक है कि ब्लॉक के प्रभाव को इस तुलना से हटा सकें।

§ २०.५ यादृच्छिकीकृत ब्लॉक अभिकल्पना के विश्लेषण के लिए एक गणितीय प्रतिरूप

मान लीजिए कि ब्लॉक i के प्रभाव को b_i से सूचित किया जाता है और j -वें किस्म के गेहूँ के प्रभाव को v_j से सूचित किया जाता है। i -वें ब्लॉक में j -वें किस्म के गेहूँ की उपज को यदि y_{ij} से सूचित किया जाता है तो

$$y_{ij} = b_i + v_j + \epsilon_{ij} \quad \dots\dots\dots(20.1)$$

यहाँ ϵ_{ij} प्लॉटों के उपजाऊपन के अंतर पर आश्रित एक यादृच्छिक चर है। पहले के उदाहरण की भाँति हम कल्पना करते हैं कि ϵ_{ij} एक प्रसामान्य चर है जिसका माध्य ० और प्रसरण σ^2 है जो ब्लॉक पर अथवा गेहूँ की किस्म पर निर्भर नहीं करते। इसके अलावा ये सब चीजें ϵ_{ij} एक दूसरे से स्वतंत्र हैं। क्योंकि हम v_j अथवा b_i का प्रयोग केवल तुलना के लिए कर रहे हैं, इसलिए हम इनको क्रमशः किस्म-प्रभाव और ब्लॉक-प्रभावों और उनके माध्यों के अंतर मान सकते हैं। इस कारण

$$\sum_{i=1}^V b_i = 0 \quad \dots\dots\dots(20.2)$$

$$\sum_{j=1}^D v_j = 0 \quad \dots\dots\dots(20.3)$$

मान लीजिए
$$\bar{Y}_j = \frac{1}{6} \sum_{i=1}^V y_{ij} \quad \dots\dots(20.4)$$

$$\begin{aligned} \text{तो} \quad \bar{y}_j &= \frac{1}{6} \sum_{i=1}^{VI} (b_j + v_j + \epsilon_{ij}) \\ &= v_j + \bar{\epsilon}_j, \quad \dots\dots\dots(20.5) \end{aligned}$$

$$\text{जहाँ} \quad \bar{\epsilon}_j = \frac{1}{6} \sum_{i=1}^{VI} \epsilon_{ij} \quad \dots\dots\dots(20.6)$$

यहाँ \bar{y}_j एक प्रसामान्य चर है जिसका माध्य v_j और प्रसरण $\frac{\sigma^2}{6}$ है। इसी प्रकार \bar{y}'_j उन प्लॉटों के प्रेक्षणों का माध्य है जिसमें j' -वी किस्म बोयी गयी है। यह भी एक प्रसामान्य चर है जिसका माध्य $v_{j'}$ और प्रसरण $\frac{\sigma^2}{6}$ है। ये दोनों चर स्वतन्त्र हैं इसलिए $(\bar{y}_j - \bar{y}'_{j'})$ भी एक प्रसामान्य चर है जिसका माध्य $(v_j - v_{j'})$ और प्रसरण $\frac{\sigma^2}{6} + \frac{\sigma^2}{6} = \frac{\sigma^2}{3}$ है। (देखिए § ८.३)

यदि निराकरणीय परिकल्पना यह है कि $v_j = v_{j'}$ तब इसके अन्तर्गत इस प्रसामान्य चर का माध्य ० होगा। यदि हमें σ^2 का मान ज्ञात हो तो इस परिकल्पना की जांच इस प्रसामान्य वचन के आधार पर कर सकते हैं। परन्तु σ^2 वास्तव में ज्ञात नहीं है और इसका अनुमान लगाने की आवश्यकता है।

§ २०.६ विभिन्न परिकल्पनाओं के अन्तर्गत σ^2 का प्राक्कलन

हम \bar{y}_j से $\frac{1}{4} \sum_{j=A}^D \bar{y}_j$ और \bar{y} से $\frac{1}{6} \sum_{i=1}^{VI} \bar{y}_i$ अथवा $\frac{1}{4} \sum_{j=A}^D \bar{y}_j$ को

सूचित करेंगे जो दोनों $\frac{1}{24} \sum_{i=1}^{VI} \sum_{j=A}^D y_{ij}$ के बराबर हैं। इसी प्रकार

$$\bar{\epsilon} = \frac{1}{4} \sum_{j=A}^D \bar{\epsilon}_j = \frac{1}{6} \sum_{i=1}^{VI} \bar{\epsilon}_i = \frac{1}{24} \sum_{i=1}^{VI} \sum_{j=A}^D \epsilon_{ij}$$

$$(1) \sum_{j=A}^D (\bar{y}_j - \bar{y})^2 = \sum_{j=A}^D \{(v_j + \bar{\epsilon}_j) - \bar{\epsilon}\}^2$$

देखिए समीकरण (20.3) और (20.5)

$$= \sum_{j=A}^D \nu_j^2 + \sum_{j=A}^D (\bar{e}_j - \bar{e})^2 \\ + 2 \sum_{j=A}^D \nu_j (\bar{e}_j - \bar{e})$$

परन्तु ν_j और \bar{e}_j स्वतंत्र हैं। इस कारण

$$\sum_{j=A}^D \nu_j (\bar{e}_j - \bar{e}) = \frac{\sum_{j=A}^D \nu_j \sum_{j=A}^D (\bar{e}_j - \bar{e})}{4} \\ = 0$$

किन्तु हर एक \bar{e}_j का वटन $N\left(0, \frac{\sigma^2}{\sqrt{6}}\right)$ है। इस कारण $\frac{\sigma^2}{6}$ का अनभिन्नत

अनुमान $\frac{1}{3} \sum_{j=A}^D (\bar{e}_j - \bar{e})^2$ है। (देखिए § १७.३.१)

यानी $\frac{1}{3} \sum_{j=A}^D \nu_j^2 + \frac{\sigma^2}{6}$ का अनभिन्नत प्राक्कलन $\frac{1}{3} \sum_{j=A}^D (\bar{y}_j - \bar{y})^2$ [है।

यदि सब ν_j बराबर हों तो

$$\nu_A = \nu_B = \nu_C = \nu_D = 0 \quad (\text{देखिए समीकरण 20.3})$$

तथा $E \left[\frac{1}{3} \sum_{j=A}^D (\bar{y}_j - \bar{y})^2 \right] = \frac{1}{6} \sigma^2 \quad \dots \dots (20.7)$

(2) इसी प्रकार

$$E \left[\frac{1}{5} \sum_{i=1}^{VI} (\bar{y}_i - \bar{y})^2 \right] = \frac{1}{5} \sum_{i=1}^{VI} b_i^2 + \frac{\sigma^2}{4} \quad \dots \dots (20.8)$$

यदि ब्लॉकों के कारण उपज पर कोई प्रभाव पड़ता हो तो

$$b_I = b_{II} = b_{III} = b_{IV} = b_V = b_{VI} = 0 \text{ और}$$

$$E \left[\frac{1}{5} \sum_{i=1}^{VI} (\bar{y}_i - \bar{y})^2 \right] = \frac{\sigma^2}{4} \quad \dots \dots (20.9)$$

§ २०.७ बिना परिकल्पना के σ^2 का प्राक्कलन

इस प्रकार हमें दो परिकल्पनाओं के अंतर्गत σ^2 के दो विभिन्न प्राक्कलक प्राप्त हुए। अब देखना यह है कि बिना परिकल्पना के भी σ^2 का अनभिन्नत प्राक्कलन सम्भव है अथवा नहीं।

$$\sum_{i=1}^{VI} \sum_{j=A}^D (y_{ij} - \bar{y})^2 = \sum_{i=1}^{VI} \sum_{j=A}^D (v_j + b_i + e_{ij} - \bar{e})^2 \text{ [दिए गए समी-}$$

करण (20.1), (20.2) और (20.3)]

$$= 6 \sum_{j=A}^D v_j^2 + 4 \sum_{i=1}^{VI} b_i^2 + \sum_{i=1}^{VI} \sum_{j=A}^D (e_{ij} - \bar{e})^2$$

$$+ 2 \sum_{i=1}^{VI} \sum_{j=A}^D v_j b_i + 2 \sum_{i=1}^{VI} \sum_{j=A}^D v_j (e_{ij} - \bar{e}) + 2 \sum_{i=1}^{VI} \sum_{j=A}^D b_i (e_{ij} - \bar{e})$$

इसमें से अंतिम तीनों राशियाँ शून्य के बराबर हैं क्योंकि v_j , b_i और e_{ij} एक दूसरे से स्वतंत्र हैं। और $E(v_j) = E(b_i) = 0$ (देखिए § ४९)

इस प्रकार $\sum_{i=1}^{VI} \sum_{j=A}^D (y_{ij} - \bar{y})^2$ का प्रत्याशित मान $6 \sum_{j=A}^D v_j^2 + 4 \sum_{i=1}^{VI} b_i^2 + 23\sigma^2$

है। हमने ये यदि $6 \sum_{j=A}^D (\bar{y}_j - \bar{y})^2 + 4 \sum_{i=1}^{VI} (\bar{y}_i - \bar{y})^2$ घटा दिया जाय तो शेष राशि का प्रत्याशित मान $15\sigma^2$ होगा। यह अनुमान किसी परिकल्पना पर आधारित नहीं है।

§ २०.८ प्रसरण विक्षेपण सारणी

इस प्रकार के कुल तीन प्राक्कलक हैं।

(1) $6 \sum_{j=A}^D (\bar{y}_j - \bar{y})^2$ यह इस परिकल्पना पर आधारित है कि गेहूँ की किस्मों

में पैदावार के दृष्टिकोण से कोई अन्तर नहीं है। या

$$v_A = v_B = v_C = v_D$$

$$(2) \quad \frac{4}{5} \sum_{i=1}^{VI} (y_i - \bar{y})^2$$

यह इस परिकल्पना पर आधारित है कि ब्लॉकों में उपजाऊपन में कोई अंतर नहीं है अथवा

$$b_I = b_{II} = b_{III} = b_{IV} = b_V = b_{VI}$$

$$(3) \quad \frac{\sum_{i=1}^{VI} \sum_{j=A}^D (y_{ij} - \bar{y})^2 - 6 \sum_{j=A}^D (\bar{y}_j - \bar{y})^2 - 4 \sum_{i=1}^{VI} (\bar{y}_i - \bar{y})^2}{15}$$

यह सभी परिवर्तनानुओं से स्वतन्त्र है। हम इन सब निष्कर्षों को एक प्रसरण-विश्लेषण सारणी के रूप में रख सकते हैं।

सारणी सरया 20 1

प्रसरण विश्लेषण सारणी

विचरण का उद्गम	वर्ग योग	स्वातन्त्र सख्या	वर्ग माध्य	वर्ग माध्य का प्रत्याक्षित मान
(1)	(2)	(3)	(4)	(5)
विलम्ब	$\sum_{i=1}^{VI} \sum_{j=A}^D (y_{ij} - \bar{y})^2 = S_1$	3	$\frac{S_1}{3} = M_1$	$\sigma^2 + \frac{6}{3} \sum_{j=A}^D \nu_j^2$
ब्लॉक	$\sum_{i=1}^{VI} \sum_{j=A}^D (y_i - \bar{y})^2 = S_2$	5	$\frac{S_2}{5} = M_2$	$\sigma^2 + \frac{4}{5} \sum_{i=1}^{VI} b_i^2$
त्रुटि	$S_e = S_e$	15	$\frac{S_e}{15} = M_e$	σ^2
कुल	$\sum_{i=1}^{VI} \sum_{j=A}^D (y_{ij} - \bar{y})^2 = S$	23		

* यह राशि S_e , S_2 और S_1 के योग को S में से घटा कर प्राप्त की जाती है। $S_e = S - (S_1 + S_2)$

§ २०९ परिकल्पनाओं की जाँच

सब ν_j शून्य के बराबर हैं इस परिकल्पना के अंतर्गत M_1 और M_2 दोनों ही σ^2 के प्राक्कलक हैं और $\frac{S_1}{\sigma^2}$ तथा $\frac{S_2}{\sigma^2}$ क्रमशः χ^2_3 और χ^2_{15} चर हैं। इस कारण $\frac{M_1}{M_2}$ $F_{3,15}$ एक चर है। (देखिए § १११) यदि प्रयोग में इस अनुपात $\frac{M_1}{M_2}$ का माप $F_{3,15}$ के पाँच प्रतिशत बिंदु से अधिक हो तो हम उस परिकल्पना को असाध्य समझेंगे जिसके आधार पर $\frac{M_1}{M_2}$ का बटन $F_{3,15}$ था अर्थात् हम इस निष्कर्ष पर पहुँचेंगे कि किस्मों में पैदावार की दृष्टि से वास्तविक अंतर है।

इसी प्रकार यदि हम यह जाँचना चाहें कि प्लॉटों के उपजाऊपन में कुछ अंतर है अथवा नहीं तो $\frac{M_2}{M_1}$ के $F_{15,3}$ चर होने का उपयोग किया जायगा। अधिकतर इस प्रकार की जाँच में वैज्ञानिक को शंका नहीं होती। यदि यह जाँच करता है तो केवल यह जानने के लिए कि प्रयोग में ब्लॉकों के निर्माण से कुछ लाभ हुआ अथवा नहीं।

यदि एक किस्मों के समान होने की परिकल्पना इस विक्षेपण द्वारा असत्य नहीं ठहरती तो अलग अलग किस्मों के युग्मों की तुलना अर्थहीन और बेकार है। परंतु यदि यह असत्य ठहराया जाता है तो हमें यह पता लगाना आवश्यक हो जाता है कि आखिर इनमें से कौन-सी किस्म सर्वात्तम है। यदि प्रेषित उपज के अनुसार इन किस्मों को क्रमबद्ध किया जाय तो दो क्रमागत (consecutive) उपजों का अंतर अर्थपूर्ण है अथवा नहीं, यह भी हम जानना चाहेंगे।

हम यह पहिले ही देख चुके हैं कि $\bar{y}_j - \bar{y}'_j$ का प्रत्याशित मान $\nu_j - \nu'_j$ है। यदि $\nu_j = \nu'_j$ हो तो $(\bar{y}_j - \bar{y}'_j)$ एक $N\left(0, \frac{\sigma^2}{\sqrt{b}}\right)$ चर होगा। इसलिए $[(\bar{y}_j - \bar{y}'_j)/M_0]\sqrt{b}$ एक t_{16} -चर होगा। इस प्रकार हम ν_j और ν'_j के बराबर होने की परिकल्पना की जाँच कर सकते हैं।

§ २०१० उदाहरण

§ २०१०१ आँकड़े

नीचे एक उदाहरण द्वारा यह सारा तरीका विस्तारपूर्वक समझाया गया है। इसी नक्के द्वारा जो पहिले दिया, विभिन्न प्लॉटों की प्रेषित पैदावार y_j दिखायी गयी है।

I

6	6
A	D
7	5
C	B

II

6	7
B	C
8	7
A	D

III

5	4
C	A
3	4
D	B

IV

3	5
C	D
8	4
A	B

V

6	4
C	D
6	4
A	B

VI

4	6
B	D
7	7
A	C

परिकलन के लिए इन आंकड़ों को नीचे दी हुई सारणी के रूप में रख दिया जाता है ।

सारणी सख्या 20-2

ब्लॉक i \ क्लस्म j	I	II	III	IV	V	VI	जोड़ $\sum_{i=1}^6 y_{ij}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A	6	8	4	8	6	7	39
B	5	6	4	4	4	4	27
C	7	7	5	3	6	7	35
D	6	7	3	5	4	6	31
जोड़ $\sum_{j=A}^D y_{ij}$	24	28	16	20	20	24	$\sum_{i=1}^4 \sum_{j=A}^D y_{ij} = 132$

§ २०.१०.२ विश्लेषण

$$\begin{aligned}
 S_1 = \text{अंतर-निस्म वर्ग-योग} &= \sum_{i=1}^{VI} \sum_{j=A}^D (\bar{y}_i - \bar{y})^2 \\
 &= \sum_{i=1}^{VI} \sum_{j=A}^D \bar{y}_i^2 - \sum_{i=1}^{VI} \sum_{j=A}^D \bar{y}^2 \\
 &= \frac{D}{j=A} \left\{ \sum_{i=1}^{VI} y_{ij} \right\}^2 - \left[\sum_{i=1}^{VI} \sum_{j=A}^D y_{ij} \right]^2 \\
 &= \frac{(39)^2 + (27)^2 + (35)^2 + (31)^2}{6} - \frac{(132)^2}{24} \\
 &= 739.33 - 726 \\
 &= 13.33
 \end{aligned}$$

$$\begin{aligned}
 S_2 = \text{अंतर-ब्लॉक वर्ग योग} &= \sum_{i=1}^{VI} \sum_{j=A}^D \bar{y}_i^2 - \sum_{i=1}^{VI} \sum_{j=A}^D \bar{y}^2 \\
 &= \frac{1}{4} [(24)^2 + (28)^2 + (16)^2 + (20)^2 + (20)^2 + (24)^2] - \frac{1}{24} (132)^2 \\
 &= 748 - 726 \\
 &= 22.00
 \end{aligned}$$

$$\begin{aligned}
 S = \text{कुल वर्ग-योग} &= \sum_{i=1}^{VI} \sum_{j=A}^D y_{ij}^2 - \sum_{i=1}^{VI} \sum_{j=A}^D \bar{y}^2 \\
 &= (6)^2 + (8)^2 + (4)^2 + (8)^2 + (6)^2 + (7)^2 \\
 &\quad + (5)^2 + (6)^2 + (4)^2 + (4)^2 + (4)^2 + (4)^2 \\
 &\quad + (7)^2 + (7)^2 + (5)^2 + (3)^2 + (6)^2 + (7)^2 \\
 &\quad + (6)^2 + (7)^2 + (3)^2 + (5)^2 + (4)^2 + (6)^2 \\
 &\quad - \frac{1}{24} (132)^2 \\
 &= 778 - 726 \\
 &= 52.00
 \end{aligned}$$

$$\begin{aligned} \therefore S_1 &= S - S_1 - S_2 \\ &= 52.00 - 13.33 - 22.00 \\ &= 16.67 \end{aligned}$$

सारणी सख्या 20-3

प्रसरण विश्लेषण सारणी

विचरण का उद्गम	वर्ग-योग	स्वातन्त्र्य सख्या	वर्ग-माध्य	अनुपात	F का 5% मान *
(1)	(2)	(3)	(4)	(5)	(6)
किस्म	$S_1 = 13.33$	5	$M_1 = 4.443$	$\frac{M_1}{M_0} = 4.03$	3.29
ब्लॉक	$S_2 = 22.00$	3	$M_2 = 4.400$	$\frac{M_2}{M_0} = 3.96$	2.90
त्रुटि	$S_3 = 16.67$	15	$M_3 = 1.110$		
कुल	$S = 52.00$	23			

इस प्रकार ब्लॉकों का अंतर और किस्मों का अंतर दोनों ही अर्थ पूर्ण हैं।

किन्ती भी ब्लॉक के प्रेक्षणों के जोड़ का प्रसरण $4\sigma^2$ होगा। इसलिए किन्ही दो ब्लॉकों के प्रेक्षण-योगों में $2 \times 1 \times \sqrt{4M_2}$ में अधिक अंतर होतो हम उसे अर्थ पूर्ण समझेंगे। (देखिए § १०६) यहाँ σ^2 का अर्थ है t_{26} का 2.5% बिंदु जिसका मान 2.131 है। (देखिए सारणी सख्या १०.१) दो ब्लॉकों के प्रभावों में वास्तविक अंतर न होने पर उनके प्रेक्षण-योगों के अन्तर के $2 \times 1 \times 1.131 \times \sqrt{4 \times 1.110} = 8.96$ से अधिक होने की प्रायिकता पाँच प्रतिशत से कम है। इस प्रकार से ब्लॉकों की तुलना के विश्लेषण को हम निम्नलिखित रूप में रख सकते हैं

II (I VI) (IV V) III

यहाँ दो ब्लॉकों को एक कोष्ठ में रखने का अर्थ है उनकी बिल्कुल समानता। ब्लॉकों को उपज के अनुसार क्रमबद्ध कर लिया गया है। ब्लॉक II में प्रेक्षित उपज सबसे आधिक है। परन्तु यह I, VI, IV अथवा V की उपज से सांख्यिकीय दृष्टिकोण

* देखिए सारणी II.1

से इतनी अधिक बड़ी नहीं है कि अंतर को अर्थपूर्ण समझा जाय। केवल II और III में अंतर सारपूर्ण समझा जा सकता है क्योंकि यह अंतर 8.96 से अधिक है। जिन ब्लॉकों में सांख्यिकीय दृष्टिकोण से अर्थपूर्ण अंतर नहीं है उनके ऊपर लिखित सकेत के अनुसार एक मोटी लकीर खीन देते हैं।

इसी प्रकार दो किस्मों के प्रेक्षणों के योगों का अंतर अर्थपूर्ण होगा यदि वह $2 \times 2.131 \times \sqrt{6 \times 110} = 11.00$ से कम न हो।

इस प्रकार $\overline{A C D B}$

अर्थात् A, C और D में कोई अर्थपूर्ण अंतर नहीं है। इसी प्रकार C, D और B में कोई अंतर नहीं है परंतु A और B का अंतर अर्थपूर्ण है। प्रेक्षणों के आधार पर उपज के अनुसार इन चार किस्मों का क्रम A, C, D और B है।

§ २०.११ ब्लॉक

यद्यपि अधिकतर प्रयोग अभिकल्पनाएँ आरम्भ में खेती के प्रयोगों के लिए ही सोच कर निकाली गयी थीं परंतु इन्हीं अभिकल्पनाओं का अन्य क्षेत्रों में भी उपयोग होता है। उदाहरण के लिए खूराक के एक प्रयोग में एक साथ पैदा हुए सूअर के बच्चों के समूह का एक ब्लॉक की तरह उपयोग किया गया था। ब्लॉक शब्द का प्रयोग अभिकल्पना में भूमि-खंड के लिए ही नहीं बल्कि किसी भी ऐसे प्रायोगिक इकाइयों के समूह के लिए किया जाता है जिसके अंदर इकाइयों को यादृच्छिकीकरण द्वारा उपचारों के साथ संयुक्त किया जाता है। इनकी संपूर्ण समष्टि को तुलना में अधिक समान (homogenous) होना चाहिए।

प्रयोग के विश्लेषण में यदि हम यह पायें कि अंतर-ब्लॉक वर्ग-माध्य और द्रुष्टि-वर्ग-माध्य का अनुपात अर्थपूर्ण है तो यह समझा जा सकता है कि ब्लॉक बनाना अभिकल्पना में लाभदायक सिद्ध हुआ है। यदि यह अनुपात अर्थपूर्ण नहीं हों तो कदाचित् यह ब्लॉक बनाना बेकार या अथवा इससे विशेष लाभ नहीं हुआ। यह ध्यान देने योग्य बात है कि यदि पिछले प्रयोग में ब्लॉक नहीं बनाये जाते तो अंतर-ब्लॉक-प्रसरण भी द्रुष्टि वर्ग माध्य में मिल जाता और यह संभव था कि किस्मों की उपज का अंतर जो इस प्रयोग के द्वारा अर्थपूर्ण ठहराया गया है—बिना ब्लॉक के प्रयोग के अर्थहीन माना जाता। इस प्रकार ब्लॉक निर्माण का प्रयोजन प्रयोग को अधिक सुगम बनाना है।

अध्याय २१

लैटिन-वर्ग अभिकल्पना (Latin Square Design)

§ २१.१ प्रयोग को सुग्राही बनाने का प्रयत्न

पिछले प्रयोग में हमने देखा था कि किसी उपचार के प्रभाव के प्राक्कलक में जो त्रुटि होती है उसका एक भाग ब्लॉकों के बीच का अंतर है। एक विशेष प्रकार की प्रयोग-अभिकल्पना द्वारा कुल त्रुटि में से इस भाग को घटाया जा सकता है और इस प्रकार प्रयोग को अधिक सुग्राही बनाया जा सकता है। यदि ब्लॉकों के अंतर के अतिरिक्त हमें त्रुटि का कोई अन्य कारण भी ज्ञात हो और उसको भी किसी विशेष अभिकल्पना द्वारा हटाया जा सके तो प्रयोग और भी अधिक सुग्राही हो जायगा। सर्वेक्षण से सबंध रखनेवाले इस प्रकार के एक प्रयोग का विवरण नीचे दिया हुआ है। इसके विश्लेषण के लिए प्रतिरूप (model) से आरंभ करके सारा सिद्धांत नहीं समझाया गया है। आशा है कि पिछले प्रयोग के प्रतिरूप और विश्लेषण को ध्यान में रखकर इसके विभिन्न चरण क्या होंगे यह आप स्वयं ही तय कर सकते हैं।

§ २१.२ उदाहरण

आजकल पंचवर्षीय योजना का बड़ा जोर है। आशा की जाती है कि पन्द्रह वर्षों में प्रति मनुष्य औसत आमदनी दुगुनी हो जायगी। भली-भाँति योजना का निर्माण करने के लिए यह जानना आवश्यक है कि भारत के निवासी इस बड़ी हुई आमदनी का प्रयोग किस प्रकार करेंगे। यद्यपि कोई भी इसकी भविष्यवाणी नहीं कर सकता, परंतु आजकल भिन्न-भिन्न आर्थिक स्थिति के लोग जिस प्रकार अपनी आमदनी खर्च करते हैं उससे इसका बहुत कुछ अनुमान हो सकता है। अब समस्या यह जानने की है कि आजकल लोग किस प्रकार खर्च करते हैं। इसके लिए सरकार की ओर से बड़े बड़े सर्वेक्षण होते हैं। इसमें कुछ मनुष्य घर-घर जाकर लोगों से उनके व्यय के विषय में पूछताछ करते हैं। आपको इस समय कल्पना करनी चाहिए कि कोई अन्वेषक आपसे आकर भिन्न-भिन्न वस्तुओं पर आपके खर्च के बारे में पूछताछ करता है। क्या आप वास्तव में उसे यह सूचना दे सकते हैं? यदि आप रोज का व्ययरेखाबंद हिस्सा रखते हैं तो आपको कुछ कठिनाई नहीं होगी। परंतु वास्तव में बहुत कम लोग ऐसे

है जो रोज का हिसाब रखते हैं। ऐसे लोगों को हिसाब केवल अनुमान से ही बताना पड़ेगा। इस दशा में त्रुटि होना प्रायः अनिवार्य है।

यद्यपि गलती को विलकुल दृष्टा देना असंभव है, परन्तु हम जानते हैं कि इस त्रुटि को दो उपादान प्रभावित करते हैं। एक तो है निर्दिष्ट काल (reference period)। यदि आप केवल पिछले दिन के खर्च के बारे में पूछें तो उसमें जितनी गलती होगी वह पिछले सप्ताह, पिछले पक्षवार अथवा पिछले माह के खर्च की जिज्ञासा के उत्तर में की हुई गलती से भिन्न होगी। इसके अलावा यह अन्वेषक पर भी निर्भर है कि वह किस प्रकार प्रश्न पूछता है। भिन्न-भिन्न प्रकार के प्रश्न पूछने से भिन्न-भिन्न प्रकार के उत्तर मिलेंगे। उदाहरण के लिए आप एक तो सीधे-सीधे यह पूछ सकते हैं कि पिछले महीने फलों पर कितना खर्चा हुआ। इसी प्रश्न को दूसरे ढंग से भी पूछा जा सकता है। अन्वेषक बारी बारी से सब फलों का नाम लेकर पूछ सकता है कि इन पर पिछले माह कितना कितना खर्च किया गया। इन सब खर्चों के जोड़ से भी महीने भर में फलों पर किये हुए खर्च का उल्लेख पता चल सकता है। एक तरीका यह भी है कि केवल फलों पर ही नहीं बल्कि अन्य वस्तुओं पर भी खर्चा पूछा जाय। इस प्रकार कुल आमदनी और खर्चों की तुलना से मासिक विभिन्न वस्तुओं पर हुए खर्चों से अधिक अच्छे अनुमान की आशा की जा सकती है।

यदि किसी मनुष्य के पास एक एक दिन का प्रत्येक फल का खर्चा लिखा हुआ है तो तीनो प्रकार से प्रश्न करने पर एक ही उत्तर मिलेगा। परन्तु उत्तर यदि याद-दास्त पर ही आधारित है तो एक ही मनुष्य इन तीन प्रकार से प्रश्न करने पर भिन्न-भिन्न उत्तर दे सकता है। इसके अलावा एक ही प्रकार के प्रश्न करने पर भी एक ही मनुष्य भिन्न-भिन्न स्थितियों में भिन्न-भिन्न उत्तर दे सकता है।

खर्चों के संबंध में हुए सर्वेक्षणों में विभिन्न निर्दिष्ट कालों और प्रश्न पूछने के भिन्न-भिन्न तरीकों का प्रयोग होता रहा है। अब प्रश्न यह उठता है कि क्या इन सर्वेक्षणों के फलों की तुलना की जा सकती है। मान लीजिए एक सर्वेक्षण उत्तर प्रदेश और एक मद्रास में होता है। क्या हम इन दो सर्वेक्षणों की मदद से यह जान सकते हैं कि मद्रास और उत्तर प्रदेश के लोगों की खर्चों की आदतें कितनी भिन्न हैं? यदि हम यह जानते हों कि इन दो सर्वेक्षणों में भिन्न-भिन्न निर्दिष्ट काल और प्रश्न पूछने के भिन्न-भिन्न तरीकों का उपयोग किया गया था, और इसके साथ यह भी जानते हों कि निर्दिष्ट काल और तरीकों के भिन्न होने से सूचना में सचमुच अंतर पड़ जाता है तो इस प्रश्न का उत्तर नकारात्मक होगा।

§ २१.३ आँकड़े

मान लीजिए, हमें चार निर्दिष्ट समयों और चार प्रश्न पूछने के तरीकों का अध्ययन करना है। इसके लिए एक प्रयोग किया जा सकता है जिसमें चार व्यक्तियों पर चारों निर्दिष्ट वालों और प्रश्न पूछने के चार तरीकों का प्रयोग करके देखा जा सकता है। यद्यपि इस प्रकार के प्रयोग में कुछ दोष है जिससे यह तुलना अभात्मक हो सकती है परन्तु इस अभिकल्पना को और उसके विश्लेषण को समझने के लिए यह उदाहरण पर्याप्त होगा।

हम उन मनुष्यों को जिन पर प्रयोग किया गया है A, B, C और D से सूचित करेंगे। प्रश्न पूछने के तरीकों को सख्याओं से और निर्दिष्ट-कालों को I, II, III और IV से सूचित किया जायगा। सारी अभिकल्पना को नीचे दिये तरीके से सारणी में रखा जा सकता है।

सारणी सत्या 21 1

निर्दिष्ट काल प्रश्न का तरीका	I	II	III	IV	कुल	माध्य
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	A 50	B 110	C 30	D 200	390	97.50
2	D 190	A 62	B 95	C 30	377	94.25
3	B 90	C 32	D 195	A 56	373	93.25
4	C 28	D 220	A 54	B 100	402	100.50
कुल	358	424	374	386	1542	
माध्य	89.50	106.00	93.50	96.50		96.38

§ २१.४ लैटिन वर्ग

इस ऊपर की सारणी में आप देखेंगे कि एक वर्ग है जिसे चार पंक्तियों (rows) और चार स्तंभों (columns) द्वारा सोलह भागों में बाँटा हुआ है। इन भागों

में चार अक्षर A, B, C और D लिखे हुए हैं। इनको इस प्रकार बाँटा गया है कि हर एक अक्षर हर एक पंक्ति और हर एक स्तंभ में एक बार और केवल एक ही बार आता है। इस प्रकार के वर्ग को लैटिन वर्ग (Latin square) कहते हैं। इस प्रयोग में एक 4×4 लैटिन वर्ग है जिसमें चार पंक्तियाँ और चार स्तंभ हैं। इसी प्रकार $5 \times 5, 6 \times 6, 7 \times 7$ इत्यादि विभिन्न परिमाणों के लैटिन वर्ग होते हैं।

§ २१.५ विश्लेषण

हर एक भाग में अक्षर के अतिरिक्त एक सख्या भी दी हुई है जो एक मास में हुए कुल खर्च को सूचित करती है। यह तीन उपादानों (factors) पर निर्भर करती है, (१) ध्यकित (२) निर्दिष्ट काल (३) प्रश्न का तरीका। इसके अलावा कुछ नुटि और रह जाती है जिसको एक प्रसामान्य चर मान कर पिछले प्रयोग की तरह विश्लेषण किया जा सकता है।

$$S_1 = \text{अंतर-निर्दिष्ट-काल वर्ग-योग} = \frac{(358)^2 + (424)^2 + (374)^2 + (386)^2}{4} - \frac{(1542)^2}{16}$$

$$= \frac{596,812}{4} - \frac{23,77,764}{16}$$

$$= 149,203 - 148,610.25$$

$$= 592.75$$

$$S_2 = \text{अंतर-प्रश्न-विधि वर्ग-योग} = \frac{(390)^2 + (377)^2 + (373)^2 + (402)^2}{4} - \frac{(1542)^2}{16}$$

$$= 148,740.50 - 148,610.25$$

$$= 130.25$$

उन सब खानों की सख्याओं का योग जिनमें A है = 222

उन सब खानों की सख्याओं का योग जिनमें B है = 395

उन सब खानों की सख्याओं का योग जिनमें C है = 120

उन सब खानों की सख्याओं का योग जिनमें D है = 805

$$\therefore S_1 = \text{यत्तर-व्यक्ति वर्ग-योग} = \frac{(223)^2 + (395)^2 + (120)^2 + (805)^2}{4}$$

$$= \frac{(1542)^2}{16}$$

$$= 216,983.50 - 148,610.25$$

$$= 68,373.25$$

$$S = \text{कुल वर्ग-योग} = [(50)^2 + (190)^2 + (90)^2 + (28)^2 + (110)^2$$

$$+ (62)^2 + (32)^2 + (220)^2 + (30)^2 + (95)^2$$

$$+ (195)^2 + (54)^2 + (200)^2 + (30)^2 + (56)^2$$

$$+ (100)^2] - \frac{(1542)^2}{16}$$

$$= 217,754.00 - 148,610.25$$

$$= 69,143.75$$

$$S_e = S - (S_1 + S_2 + S_3) = 69,143.75 - (592.75 + 130.25 + 68,373.25)$$

$$= 47.50$$

इन सब परिवर्तनों को प्रसरण-विश्लेषण सारणी के रूप में रखा जा सकता है।

सारणी सख्या 212

संतिन-वर्ग अभिकल्पना के लिए प्रसरण-विश्लेषण

विचरण का उदगम	स्वातन्त्र्य सख्या	वर्ग-योग	वर्ग माध्य	अनुपात	F का 5% मान
(1)	(2)	(3)	(4)	(5)	(6)
निर्दिष्ट समय	3	$S_1 = 592.75$	$M_1 = 197.58$	$\frac{M_1}{M_c} = 24.95$	4.76
प्रश्न विधि	3	$S_2 = 130.25$	$M_2 = 43.42$	$\frac{M_2}{M_c} = 5.48$	4.76
व्यक्ति	3	$S_3 = 68,373.25$			
त्रुटि	6	$S_e = 47.50$	$M_c = 7.92$		
कुल	15	$S = 69,143.75$			

निर्दिष्ट काल और प्रश्न विधि दोनों के लिए प्रसरण अनुपात अर्थपूर्ण है क्योंकि $F_{3,6}$ का पाँच प्रतिशत बिंदु 4.76 है। (देखिए सारणी सख्या 11.1) वास्तव में निर्दिष्ट काल के लिए अनुपात तो बहुत अधिक अर्थ-पूर्ण है क्योंकि यह $F_{3,6}$ के 0.1 प्रतिशत बिंदु 23.70 से भी अधिक है। इस कारण अब हम प्रश्न के तरीकों के युग्मों और निर्दिष्ट कालों के युग्मों की तुलना करना चाहेंगे।

यदि हम दो निर्दिष्ट कालों की तुलना करना चाहें तो इसके लिए हमें उन दोनों निर्दिष्ट कालों के लिए जो माध्य है उनका अंतर लेना होगा। क्योंकि ये दोनों माध्य चार चार प्रेक्षणों पर आधारित हैं इस कारण इनके प्रसरण $\frac{\sigma^2}{4}$ हैं जहाँ σ^2 एक अकेले प्रेक्षण का प्रसरण है। इस कारण इनके अंतर का प्रसरण $\frac{\sigma^2}{4} + \frac{\sigma^2}{4} = \frac{\sigma^2}{2}$ है।

यदि इनके अंतर X का माध्य शून्य हो तो $\frac{X}{\sigma/\sqrt{2}}$ एक प्रसामान्य $N(0,1)$ चर समझा जा सकता है। इसके अतिरिक्त (तुट्टि वर्ग योग) $\div \sigma^2$ एक χ^2_6 चर है इस कारण $\frac{X}{\sigma/\sqrt{2}} \div \frac{\sqrt{\text{तुट्टि वर्ग माध्य}}}{\sigma}$ एक t_6 चर होगा।

यदि t_6 के पाँच प्रतिशत बिंदु को t से सूचित किया जाय तो X का मान $t \frac{\sqrt{\text{तुट्टि वर्ग माध्य}}}{2}$ से अधिक होने पर हमें X के माध्य के शून्य होने में संदेह होगा।

ऊपर की सारणी (21.2) में तुट्टि-वर्ग-माध्य = 7.92 है। अतः $t \frac{\sqrt{\text{तुट्टि वर्ग-माध्य}}}{2}$

$$= 2.447 \times \sqrt{\frac{7.92}{2}} = 4.87$$

इस प्रकार निर्दिष्ट कालों के लिए

II IV III I

तथा तरीकों के लिए

4 1 2 3

(देखिए सारणी सख्या 21.1)

§ २१.६ साधारण

लैटिन वर्ग अभिकल्पना का प्रयोग खेती सबधी प्रयोगों में अधिक होता है। उसमें वास्तव में धरती पर ग्रह वर्ग बनाया जाता है और पौधों की जिन किस्मों की तुलना करनी हो उन्हें इस प्रकार लगाया जाता है कि हर एक पक्ति और हर एक स्तम्भ में एक किस्म केवल एक ही बार बोयी जाय। इस प्रकार के प्रयोग का विश्लेषण उदाहरण में दिये हुए ढग से किया जाता है। ऊपर के प्रयोग में यदि यादृच्छिकीयत बलॉरु-अभिकल्पना का उपयोग किया जाता तो हम एक प्रयोग द्वारा केवल एक ही परिकल्पना की जाँच कर सकते थे—तरीके से सवधित अथवा निर्दिष्ट काल से सवधित। परन्तु लैटिन-वर्ग के रूप में रखने से इन दोनों ही परिकल्पनाओं की जाँच एक ही प्रयोग के विश्लेषण की सहायता से की जा सकती है।

खेती सबधी प्रयोगों में इसका उद्देश्य किस्मों के प्रभाव को दो अलग-अलग प्रभावों, पक्ति-प्रभाव और स्तम्भ प्रभाव से मुक्त करना होता है। उनमें हम केवल एक ही परिकल्पना की जाँच करना चाहते हैं—बहु यह कि किस्मों में कोई विशेष अन्तर नहीं है। इस प्रकार आपने देखा कि इस प्रयोग-अभिकल्पना का अलग-अलग उद्देश्यों से उपयोग किया जा सकता है, परन्तु विश्लेषण की विधि वही रहती है।

अध्याय २२

बहु-उपादानीय प्रयोग

(Factorial Experiments)

§ २२.१ परिचय

अब तक आप यह समझ ही गये होंगे कि किसी प्रयोग को अनेक उपादान प्रभावित कर सकते हैं। यदि ये प्रभाव सयोग्य (additive) हों तो हम इनको एक एक करके माप सकते हैं। ऊपर के प्रयोग में यदि हम केवल एक ही व्यक्ति से एक ही निश्चित काल के सबंध में विभिन्न तरीकों से प्रश्न करते तो उसमें उत्तर प्रदानता केवल तरीकों के भिन्न होने के कारण आता और मौसल अंतर के धुन्य-प्राय होने की परिकल्पना की जाँच की जा सकती थी। इसी सिद्धान्त का उपयोग यादृच्छिकीकृत प्लॉट अभिकल्पना में भी किया जाता है। परन्तु दो उपादानों के महत्त्वपूर्ण होने पर इस प्रकार के प्रयोग अर्थात् हो जाते हैं और हमें लैटिन वर्ग इत्यादि अन्य प्रयोग-अभिकल्पनाओं की शरण लेनी पड़ती है। परन्तु इनका विश्लेषण उस दशा में ही सतोपजनक हो सकता है जब इनके प्रभाव सयोग्य हों। ऊपर के उदाहरण में यह संभव है कि विशेष प्रश्नविधि का प्रभाव विभिन्न व्यक्तियों पर अलग-अलग हो। यह भी हो सकता है कि विभिन्न प्रश्नविधियों के संयोजन में एक ही निश्चित काल का अलग-अलग प्रभाव पड़ता हो। ऐसी दशा में जब उपादानों का प्रभाव सयोग्य न हो तो सब एक ही प्रश्न के लिए यह पृथक प्रयोग करना सम्भव नहीं है। या तो प्रयोग से किमी प्रश्न का भी सतोपजनक उत्तर नहीं मिलेगा अथवा कई उपादानों के बिषय में बहुत से प्रश्नों का उत्तर एक साथ ही मिल जायगा।

यद्यपि हम कुछ विशेष उपादानों का अध्ययन करना अधिक उपयुक्त और आवश्यक समझते हों तथापि बहुधा यह कहना कठिन होता है कि इनमें से सबसे अधिक महत्त्वपूर्ण कौन-सा है। हमें पहले से यह ज्ञात होना भी संभव नहीं कि एक उपादान का प्रभाव दूसरे उपादानों के प्रभाव से सर्वथा स्वतंत्र है अथवा नहीं। जब कुछ विशेष उपादानों को प्रयोग के लिए चुना जाता है तो उसका कारण यह नहीं होता कि वे ही सबसे अधिक महत्त्वपूर्ण हैं बल्कि केवल यह कि इन उपादानों पर अधिक

आसानी से नियंत्रण किया जा सकता है और इनको सरलता से नापा जा सकता है। कोई भी जटिल मशीन अथवा औद्योगिक प्रणाली अवश्य ही अन्य उपादानों से भी प्रभावित होती होगी। मजदूर, मशीन तथा कच्चा माल तीनों में से किसी भी एक का प्रभाव अन्य दोनों उपादानों के प्रभावों से जुड़ा हो सकता है। दो उपादानों के इस प्रकार एक दूसरे से प्रभावित होने को परस्पर-क्रिया (interaction) कहते हैं। किसी भी उपादान के प्रभाव को पूर्णरूप से समझने के लिए यह आवश्यक है कि अन्य उपादानों से उसकी परस्पर-क्रिया का भी ज्ञान हो। यदि उपादानों के लिए अलग-अलग जाँच होती है तो इसका कारण यह नहीं है कि इस प्रकार अलग जाँच करना उपयुक्त वैज्ञानिक रीति है। बहुधा गलती से यह मान लिया जाता है कि एक साथ सब उपादानों पर प्रयोग करना असुविधाजनक है किन्तु यह बात सच नहीं है।

हम नीचे खेती सबधी एक बहु-उपादानीय प्रयोग का वर्णन करेंगे जिससे हमें यह पता चलेगा कि एक साथ अनेक उपादानों के मुख्य प्रभाव (main effect) और उनकी परस्पर-क्रियाओं (interactions) को किस प्रकार नापा जाता है, और कैसे उनके धून्य-प्राय होने की परिवर्तना की जाँच की जाती है।

§ २२.२ बहु-उपादानीय प्रयोग के लाभ

एक नये किस्म के चावल की विदेशों में बहुत चर्चा है और उसे भारत में प्रवेश कराने की योजना बनायी जा रही है। आजकल जिन किस्मों के चावल भारत में बोये जाते हैं उनसे यह किस्म वास्तव में श्रेष्ठ है अथवा नहीं, यह निश्चय रूप से नहीं कहा जा सकता। यहाँ श्रेष्ठता स्वाद से नहीं बल्कि पैदावार के दृष्टिकोण से मापी जा रही है क्यों कि इस समय सबसे बड़ी समस्या अन्न-संकट को टालना है। इसके अतिरिक्त यह भी पता नहीं कि चावल को बोने, उसमें जल देने और देखभाल करने आदि की सर्वश्रेष्ठ विधि क्या है। किस किस्म की खाद कितनी मात्रा में देना सर्वोत्तम होगा, यह भी खोज कर पता लगाने की बात है। यह हो सकता है कि कोई खाद किसी किस्म के चावल के लिए और कोई अन्य खाद किसी दूसरी किस्म के चावल के लिए उपयुक्त हो। यह भी हो सकता है कि पौधों को दूर-दूर बोने पर जो किस्म सबसे अधिक उपज देती है वही पौधों को पास पास बोने पर निकृष्ट सिद्ध हो।

ऐसी दशा में बोने की किसी विशेष रीति और खाद को लेकर यदि किस्मों की तुलना की जाय तो यह अमाल्गम होगी। यह सम्भव है कि उपादानों में परस्पर क्रिया न हो। उपादानों, किस्म, बोने की रीति और खाद के प्रभाव वास्तव में संयोज्य हो।

परन्तु फिर भी एक बहुउपादानीय प्रयोग के मुकाबले में अलग-अलग उपादानों के लिए अलग-अलग प्रयोग करना कम दक्ष (efficient) है। इसका कारण यह है कि बहु-उपादानीय प्रयोग में एक ही प्लॉट का अलग-अलग उपादानों के प्रभाव को आंकने के लिए अनेक बार उपयोग करना होता है।

उदाहरण के लिए मान लीजिए कि एक प्रयोग में तीन उपादान हैं, जिनमें से प्रत्येक के दो-दो स्तर (level) हैं। इस प्रकार कुल $2 \times 2 \times 2 = 8$ सचय इन उपादानों के स्तरों के होंगे। यदि प्रत्येक सचय का पांच बार प्रयोग किया जाय तो कुल $8 \times 5 = 40$ प्लॉटों की आवश्यकता होगी। किसी भी एक उपादान के मुख्य प्रभाव के लिए उन 20 प्लॉटों के प्रेषणों के माध्य की तुलना जिनमें यह उपादान एक विशेष स्तर पर है, उन अन्य 20 प्लॉटों के प्रेषणों के माध्य से की जायगी जिनमें यह दूसरे स्तर पर है। यदि हम अलग-अलग उपादानों के लिए अलग-अलग प्रयोग करें जिनमें मुख्य प्रभाव का इसी प्रकार 20 प्लॉटों के माध्यों के अंतर द्वारा प्राक्कलन किया जाय तो कुल $40 \times 3 = 120$ प्लॉटों की आवश्यकता होगी। यही कार्य एक बहुउपादानीय प्रयोग में केवल 40 प्लॉटों द्वारा सगम होता है।

§ २२३ मुख्य प्रभाव और परस्पर-क्रिया

विभिन्न स्तरों पर दूसरे उपादानों के सहयोग से उत्पन्न किसी एक उपादान के प्रभावों के माध्य को इस उपादान का मुख्य प्रभाव (main effect) कहते हैं। ऊपर के उदाहरण में मान लीजिए कि दो किस्में V_1 और V_2 दो बोनो के तरीके S_1 और S_2 और दो खाद M_1 और M_2 हैं। ये तीनों उपादान दो-दो स्तरों पर हैं। इन उपादानों के निम्नलिखित $2^3 = 8$ सचय हो सकते हैं।

- (1) $V_2 S_1 M_1$ (2) $V_2 S_1 M_2$ (3) $V_2 S_2 M_1$ (4) $V_2 S_2 M_2$
 (5) $V_1 S_1 M_1$ (6) $V_1 S_1 M_2$ (7) $V_1 S_2 M_1$ (8) $V_1 S_2 M_2$

यदि इन आठ सचयों को एक ब्लॉक के आठ प्लॉटों में यादृच्छिकीकरण द्वारा बाँटा जाय तो होनेवाली पैदावार इन सचयों के प्रभाव और प्लॉटों के प्रभाव का योग होगी। यादृच्छिकीकरण के कारण प्लॉटों का प्रभाव प्रत्येक सचय के लिए समान है। हमारे प्रतिरूप के अनुसार यह प्रभाव ϵ एक $N(0, \sigma)$ चर है। मान लीजिए एक ब्लॉक के जिस प्लॉट में $V_2 S_1 M_1$ का उपयोग हुआ है उसकी उपज ($V_2 S_1 M_1$) है और जिस प्लॉट में $V_1 S_2 M_2$ का उपयोग हुआ है उसकी उपज ($V_1 S_2 M_2$) है।

इसलिए इन दो सचयों के प्रभावों के अंतर का प्राक्कलन $= (V_2 S_1 M_1) - (V_1 S_1 M_1)$ परंतु इन दोनों सचयों में बौने के तरीके और खादें समान हैं। इसलिए इन सचयों के प्रभाव के अंतर को किस्मों का प्रभाव समझा जा सकता है। क्योंकि यह प्रभाव अन्य दोनों उपादानों के स्तर पर भी निर्भर कर सकता है इसलिए इस प्रभाव को $V | S_1 M_1$ से सूचित किया जायगा। इसी प्रकार हम $V | S_1 M_2$, $V | S_2 M_1$ तथा $V | S_2 M_2$ की परिभाषा कर सकते हैं। किस्म के इन चार प्रभावों के माध्य को जो अन्य उपादानों के विभिन्न स्तरों पर होते हैं, हम किस्म का मुख्य प्रभाव कहते हैं और इसे V से सूचित करते हैं।

इस तरह V का अनभिन्नत प्राक्कलन \hat{V} निम्नलिखित है

$$\begin{aligned} \hat{V} &= \frac{1}{4} \left[\{(V_2 S_1 M_1) - (V_1 S_1 M_1)\} + \{(V_2 S_1 M_2) - (V_1 S_1 M_2)\} \right. \\ &\quad \left. + \{(V_2 S_2 M_1) - (V_1 S_2 M_1)\} + \{(V_2 S_2 M_2) - (V_1 S_2 M_2)\} \right] \\ &= \frac{1}{4} (V_2 - V_1) (S_2 + S_1) (M_2 + M_1) \quad \dots (22.1) \end{aligned}$$

इस प्रकार उन सब प्लॉटों की पैदावारों के योग में से जिनमें V_2 का प्रयोग हुआ है अन्य प्लॉटों की पैदावारों के योग को घटाने और कुल उन प्लॉटों की जिनमें V_1 बोया गया है सख्या से विभाजित करने पर हमें V_2 और V_1 के प्रभावों के औसत अंतर V का प्राक्कलन प्राप्त होता है।

इसी प्रकार अन्य उपादानों के मुख्य प्रभावों की परिभाषा की जा सकती है।

$$\hat{S} = \frac{1}{4} (V_2 + V_1) (S_2 - S_1) (M_2 + M_1) \quad \dots (22.2)$$

$$\hat{M} = \frac{1}{4} (V_2 + V_1) (S_2 + S_1) (M_2 - M_1) \quad \dots (22.3)$$

यदि $(V_1 S_1 M_1)$ इत्यादि एक ब्लॉक के एक प्लॉट की पैदावार नहीं बल्कि b प्लॉटों के एक एक प्लॉट यानी कुल b प्लॉटों की पैदावार का माध्य हो तो इनमें से प्रत्येक का प्रसरण $\frac{\sigma^2}{b}$ तथा ऊपर के तीनों प्राक्कलनों के प्रसरण

$$\frac{1}{(4)^2} \times 8 \frac{\sigma^2}{b} = \frac{\sigma^2}{2b} \text{ है।}$$

मान लीजिए, हम $V | S_2 M_1$ के प्राक्कलक में से $V | S_1 M_1$ के प्राक्कलक को घटाते हैं। यह $S_2 M_1$ तथा $S_1 M_1$ पर V के प्रभावों के अंतर का प्राक्कलक होगा।

इस अंतर से यह पता चलता है कि खाद का स्तर M_1 होने पर बोने की विधि का किस्म के प्रभाव पर क्या असर पड़ता है। इसी प्रकार खाद का स्तर M_2 दिया होने पर हम एक अन्य अंतर को प्राप्त कर सकते हैं। इन दो अंतरों के माध्य को दो से विभाजित करने पर हमें जो राशि मिलती है उसे हम किस्म और बोने की विधि की परस्पर-क्रिया (interaction) VS का प्राक्कलक कहते हैं। इस प्रकार

$$\begin{aligned} \widehat{VS} &= \frac{1}{4} \left[\{(V_2 S_2 M_1) - (V_1 S_2 M_1)\} - \{(V_2 S_1 M_1) - (V_1 S_1 M_1)\} \right. \\ &\quad \left. + \{(V_2 S_2 M_2) - (V_1 S_2 M_2)\} - \{(V_2 S_1 M_2) - (V_1 S_1 M_2)\} \right] \\ &= \frac{1}{4} (V_2 - V_1) (S_2 - S_1) (M_2 + M_1) \quad \dots (22.4) \end{aligned}$$

इसी प्रकार VM और MS के प्राक्कलक निम्नलिखित होंगे

$$\widehat{VM} = \frac{1}{4} (V_2 - V_1) (S_2 + S_1) (M_2 - M_1) \quad \dots \dots (22.5)$$

$$\widehat{SM} = \frac{1}{4} (V_2 + V_1) (S_2 - S_1) (M_2 - M_1) \quad \dots \dots (22.6)$$

ये तीनों द्वि-उपादानीय परस्पर-क्रियाएँ हैं क्योंकि इनमें केवल दो उपादानों के एक दूसरे पर प्रभाव का मिश्रण किया गया है। यदि हम खाद का स्तर M_2 दिये होने पर किस्म और बोने की विधि की परस्पर-क्रिया

$$\widehat{VS} | M_2 = \frac{1}{2} (V_2 - V_1) (S_2 - S_1) M_2$$

तथा खाद के स्तर M_1 पर किस्म और बोने की विधि की परस्पर-क्रिया

$$\widehat{VS} | M_1 = \frac{1}{2} (V_2 - V_1) (S_2 - S_1) M_1$$

के अंतर को लें तो यह किस्म और बोनो की विधि की परस्पर-क्रिया पर खाद के प्रभाव का प्राक्कलक है। इस अंतर को दो से विभाजित करने पर हमें त्रि-उपादानीय परस्पर क्रिया VMS का प्राक्कलक प्राप्त होता है।

$$\widehat{VMS} = \frac{1}{4} (V_2 - V_1) (M_2 - M_1) (S_2 - S_1) \quad \dots (227)$$

यह ध्यान देने की बात है कि परस्पर-क्रियाओं के उपादानों का क्रमचय (permutation) करने से कोई अंतर नहीं पड़ता। उदाहरण के लिए $VS=SV$ अथवा $VMS=VSM=MVS$ इत्यादि। इसके अतिरिक्त मुख्य प्रभावों और परस्पर-क्रियाओं की परिभाषा इस प्रकार दी गयी है कि इन सबके प्रसरण $\frac{\sigma^2}{2b}$ हैं।

§ २२४ उदाहरण

अब आप यह तो समझ गये होंगे कि मुख्य प्रभावों और परस्पर-क्रियाओं का अनुमान किस प्रकार किया जा सकता है। सैनी सबधी प्रयोगों का मुख्य उद्देश्य भी यही होता है। परन्तु इसके अलावा हम कुछ निराकरणिय परिकल्पनाओं की जाँच भी करना चाहेंगे जिनका तात्पर्य यह जानना है कि मुख्य प्रभाव आदि के अनुमानों का शून्य से जो अंतर है वह अर्थपूर्ण है अथवा नहीं। इन परिकल्पनाओं की जाँच के बाद हम उपादान-सचयों को उत्कृष्टता के क्रम में रख सकेंगे।

इस ऊपर लिखित प्रयोग में कुल आठ उपचार हैं। इन सबको एक ब्लॉक के आठ प्लॉटों में यादृच्छिकीकरण द्वारा बाँटा जा सकता है। इस प्रकार के कई ब्लॉक लेने से हमें एक यादृच्छिकीकृत ब्लॉक-अभिकल्पना प्राप्त होती है। इसका विश्लेषण किस प्रकार किया जा सकता है, यह तो आप जानते ही हैं। परन्तु अंतर उपचार वर्ग-योग को हम फिर मुख्य प्रभावों और परस्पर क्रियाओं से संबंधित वर्ग-योगों में विभाजित कर सकते हैं और इनमें से प्रत्येक को F -परीक्षण द्वारा जाँचा जा सकता है। मुख्य प्रभावों और परस्पर-क्रियाओं की स्वातंत्र्य-सहायता केवल एक एक होने के कारण इनको t -परीक्षण द्वारा जाँचना अधिक सरल है। यह सब किस प्रकार किया जायगा, वह निम्नलिखित उदाहरण द्वारा स्पष्ट हो जायगा।

सारणी संख्या 22.1

बहु-उपादानीय प्रयोग के आंकड़े

ब्लॉक	I	II	III	IV	कुल	माध्य
उपचार	(2)	(3)	(4)	(5)	(6)	(7)
$V_1 M_1 S_1$	3	5	4	4	16	4
$V_2 M_1 S_2$	5	6	5	4	20	5
$V_1 S_2 M_1$	6	8	5	5	24	6
$V_2 S_2 M_1$	7	10	8	7	32	8
$V_1 S_1 M_2$	5	7	7	5	24	6
$V_2 S_1 M_2$	6	8	6	4	24	6
$V_1 S_2 M_2$	10	12	10	8	40	10
$V_2 S_2 M_2$	14	15	11	8	48	12
कुल	56	71	56	45	228	
माध्य	7 000	8 875	7 000	5 625		7 125

§ २२.५ विश्लेषण

$$\begin{aligned}
 \text{ब्लॉक वर्ग-योग } S_1 &= (56 \times 7\,000) + (71 \times 8\,875) + (56 \times 7\,000) \\
 &\quad + (45 \times 5\,625) - (228 \times 7\,125) \\
 &= (392\,000 + 630\,125 + 392\,000 + 253\,125) - 1624\,500 \\
 &= 42\,750
 \end{aligned}$$

$$\text{उपचार वर्ग-योग } S_2 = (16 \times 4) + (20 \times 5) + (24 \times 6) + (32 \times 8)$$

$$\begin{aligned}
 &+(24 \times 6) + (24 \times 6) + (40 \times 10) + (48 \times 12) \\
 &-(228 \times 7 \ 125) \\
 &= 64 + 100 + 144 + 256 + 144 + 400 \\
 &+ 576 - 1624 \ 500 \\
 &= 1828 \ 000 - 1624 \ 500 \\
 &= 203 \ 500
 \end{aligned}$$

$$\begin{aligned}
 \text{कुल वग योग } S &= (3)^2 + (5)^2 + (4)^2 + (4)^2 + (5)^2 + (6)^2 + (5)^2 + (4)^2 \\
 &+ (6)^2 + (8)^2 + (5)^2 + (5)^2 + (7)^2 + (10)^2 + (8)^2 + (7)^2 \\
 &+ (5)^2 + (7)^2 + (7)^2 + (5)^2 + (6)^2 + (8)^2 + (6)^2 + (4)^2 \\
 &+ (10)^2 + (12)^2 + (10)^2 + (8)^2 + (14)^2 + (15)^2 + (11)^2 + (8)^2 \\
 &-(228 \times 7 \ 125) \\
 &= 1894 \ 000 - 1624 \ 500 \\
 &= 269 \ 500
 \end{aligned}$$

$$\begin{aligned}
 \text{त्रुटि वग-योग } Se &= S - S_1 - S_2 \\
 &= 269 \ 500 - 42 \ 750 - 203 \ 500 \\
 &= 23 \ 250
 \end{aligned}$$

सारणी सख्या 22.2

प्रसरण विश्लेषण सारणी

विचरण का उद्गम	स्वातंत्र्य सख्या	वग-योग	वग माध्य	अनुपात	5% स्तर पर अधपूण मान*
(1)	(2)	(3)	(4)	(5)	(6)
ब्लॉक	3	$S_1 = 42 \ 75$	$M_1 = 14 \ 25$	$\frac{M_1}{Me} = 12 \ 84$	3 07
उपचार	7	$S_2 = 203 \ 50$	$M_2 = 29 \ 07$	$\frac{M_2}{Me} = 26 \ 19$	2 48
त्रुटि	21	$Se = 23 \ 25$	$Me = 1 \ 11$		
कुल	31	$S = 269 \ 50$			

* * (देखिए सारणी सख्या 111)

इन प्रकार हम देखते हैं कि उपचार और ब्लॉक दोनों के वर्ग-योग अर्थपूर्ण हैं। वास्तव में ये पाँच प्रतिशत स्तर पर ही नहीं बल्कि ०.१% स्तर पर भी अर्थपूर्ण हैं।

अब हम उपादानों के मुख्य प्रभाव तथा परस्पर-क्रियाओं का परिकलन निम्न-लिखित सारणी को सहायता से करते हैं।

सारणी संख्या 22 3

उपादानों के प्रभावों का परिकलन करने के लिए सारणी

उपचार	उपज	(1)	(2)	(3)	मुख्य प्रभाव, परस्पर-क्रिया
(1)	(2)	(3)	(4)	(5)	(6)
$V_1 S_1 M_1$	4)	9)	23)	57	सब प्रभावों का योग
$V_2 S_1 M_1$	5)	14)	34)	5	
$V_1 S_2 M_1$	6)	12)	3)	45	4S
$V_2 S_2 M_1$	8)	22)	2)	3	4VS
$V_1 S_1 M_2$	6)	1)	5)	11	4M
$V_2 S_1 M_2$	6)	2)	10)	-1	4VM
$V_1 S_2 M_2$	10)	0)	1)	5	4SM
$V_2 S_2 M_2$	12)	2)	2)	1	4VSM

ऊपर की सारणी में मुख्य प्रभावों और परस्पर-क्रियाओं का परिकलन करने की सरल रीति दी हुई है। प्रथम कालम में उपचारों के नाम दे रते हैं। इनको एक विशेष क्रम में सजाया गया है। पहिले $V_1 S_1 M_1$ जिसमें प्रत्येक सूचनाक (index) 1 है। उसके बाद $V_2 S_1 M_1$ जिसमें केवल V का सूचनाक 2 है। उसके पश्चात् $V_1 S_2 M_1$ जिसमें केवल S का सूचनाक 2 है। इसके बाद $V_2 S_2 M_1$ है जिसमें V और S दोनों के सूचनाक 2 हैं। इस प्रकार V और S के अकेले और साथ-साथ सूचनाक 2 पाने के बाद M की बारी आती है और अकेले उसका सूचनाक 2 रखा जाता है। उसके पश्चात् क्रमशः V और M; S और M तथा V, S और M को सूचनाक 2 दिये गये हैं।

दूसरे स्तंभ में दान उपचारों के लिए माध्य उपज दी हुई है जिनका परिकलन पहिले ही किया जा चुका है - (सारणी 22 1)। इनको दो दो के युग्मों में बाँट दिया गया है। तीसरे स्तंभ में पहिली चार सख्याएँ क्रमशः इन युग्मों के जोड़ों से और अंतिम चार

सख्याएँ इन युग्मों के अतरो से बनी हैं। इन सख्याओं को फिर दो-दो के युग्मों में बाँट दिया गया है। चौथे स्तम्भ में फिर वही क्रिया दुहरायी गयी है। यानी प्रथम चार सख्याएँ क्रमशः तीसरे स्तम्भ में दिये हुए युग्मों के जोड़ों से और अन्य चार इनके अतरो से बनी हैं। इस क्रिया को अंतिम बार पाँचवें स्तम्भ में दुहराया गया है। इस स्तम्भ की सख्याएँ मुख्य प्रभाव और परस्पर क्रियाएँ हैं जैसा कि 22.1 से 22.7 सत्यक समीकरणों से प्रकट है। जिन प्रभावों के ये अनुमान हैं उन्हें छठे स्तम्भ में दिया गया है। आपने यह नोट किया होगा कि उपचार में जिन जिन एक, दो, या तीन उपादानों के सूचकांक 2 हैं उनके सामने उन्ही उपादानों के संयुक्त प्रभावों का अनुमान दे रखा है।

क्योंकि मुख्य प्रभावों और परस्पर-क्रियाओं की कुल सख्या 7 है और 1—परीक्षण के लिए प्रत्येक को त्रुटि-वर्ग-माध्य के वर्ग मूल से विभाजित किया जायगा इसलिए बजाय प्रत्येक प्रभाव के लिए 1 के मान का परिकलन करने के यह मालूम करना अधिक सरल होगा कि वह मान क्या है जिससे अधिक होने पर इनमें से किसी को भी अर्थपूर्ण समझा जा सके।

स्वातन्त्र्य सख्या 21 के लिए 1—बटन का पाँच प्रतिशत बिंदु 2.08 है (देखिए सारणी सख्या 10.1)। इन सब प्रभावों के प्राक्कलनों का प्रसरण $\frac{\sigma^2}{8}$ है। पाँचवें स्तम्भ में दी हुई सख्याओं का प्रसरण $2\sigma^2$ है। इसलिए यदि इस स्तम्भ की कोई सख्या $2.08 \times \sqrt{2(\text{त्रुटि वर्ग-माध्य})}$ से अधिक हो तो वह अर्थपूर्ण है। (देखिए § १०.२) यहाँ $2.08 \times \sqrt{2(\text{त्रुटि वर्ग-माध्य})} = 3.10$

इस प्रकार हम देखते हैं कि V, S, M तथा SM अर्थपूर्ण हैं। किस्म V_1 से किस्म V_2 अधिक उपज देती है चाहे उसके साथ किसी भी बोने की विधि और खाद का प्रयोग किया जाय। इसी प्रकार S_1 से S_2 अच्छी बोने की विधि है और M_1 से M_2 अच्छी खाद है। परंतु S_2 और M_2 का संयुक्त प्रभाव उन दोनों के अलग-अलग प्रभावों के योग से भी अधिक है क्योंकि SM का प्राक्कलन घनात्मक है। इससे यह पता चलता है कि सर्वोत्तम उपचार $V_2 S_2 M_2$ है।

यह हम पहले ही कह चुके हैं कि मुख्य प्रभावों और परस्पर-क्रियाओं के वर्गों के योग उपचार वर्ग-योग के बराबर है। इस जवाहरण में हम इन कथन की जाँच कर सकते हैं। हमें देखना है कि (सारणी सख्या 22.3 के अनुसार)

$$\frac{(5)^2 + (15)^2 + (3)^2 + (11)^2 + (-1)^2 + (5)^2 + (1)^2}{2} = \text{उपचार-वर्ग योग}$$

$$\text{अथवा } \frac{25 + 225 + 9 + 121 + 1 + 25 + 1}{2} = \text{उपचार-वर्ग योग}$$

$$\text{अथवा } \frac{407}{2} = 203.5 = \text{उपचार-वर्ग योग}$$

यह उपचार वर्ग-योग का वही मान है जिसका परिकल्पन उपचार-योगी द्वारा करके हमने प्रसरण विश्लेषण सारणी में रखा था ।

समाकुलन (Confounding)

§ २३ १ असंपूर्ण-ब्लॉक अभिकल्पना की आवश्यकता

अभी तक हमने जितनी भी अभिकल्पनाओं का अध्ययन किया है उनमें जितने भी उपचार (treatments) थे उन सबको प्रत्येक ब्लॉक में शामिल किया गया था। आपको याद होगा कि ब्लॉक बनाने का उद्देश्य यह था कि एक ही ब्लॉक में जो प्लॉट हो उनमें विशेष अंतर न हो। यदि प्लॉटों की संख्या बहुत अधिक न हो तो ब्लॉक में इस प्रकार की समागता (homogeneity) होना बहुत कठिन नहीं है। कृपि सबधी प्रयोगों में पास के प्लॉटों में अधिक अंतर नहीं होता। परंतु यदि दस दस या बारह बारह प्लॉट एक एक ब्लॉक में हो तो दो छोरों के प्लॉटों में काफी अंतर हो सकता है। यदि अंतर अधिक हो तो ब्लॉक बनाना व्यर्थ हो जाय। इस कारण उपचारों की संख्या अधिक हो जाने पर हमें अन्य अभिकल्पनाओं की तलाश करनी पड़ती है।

इन अभिकल्पनाओं को हम असंपूर्ण-ब्लॉक अभिकल्पना (incomplete block design) की संज्ञा देते हैं। इनमें ब्लॉक के प्लॉटों की संख्या कुल उपचारों की संख्या से कम होती है। यदि प्रयोग-बहु-उपादानिय हो तो हम इस प्रकार के प्रयोग द्वारा सभी मुख्य प्रभावों और परस्पर-क्रियाओं का अनुमान नहीं लगा सकते। इस दशा में हमें यह सोचना पड़ता है कि कौन से मुख्य प्रभाव या परस्पर-क्रियाएँ सबसे कम महत्व रखती हैं। प्रयोग-अभिकल्पना इस प्रकार बनायी जाती है कि इन महत्वहीन प्रभावों को छोड़कर अन्य सब का अनुमान हम लगा सकें और अन्य प्रभावों से संबंधित निराकरणिय परिवर्तनाओं की हम जाँच कर सकें। यह देखा गया है कि अधिकतर उच्च-क्रम (higher order) की परस्पर-क्रियाएँ महत्वपूर्ण नहीं होती और इन्हीं का हमें दलितान करना पड़ता है। जब हम किसी प्रभाव का अनुमान नहीं लगा सकते और न यह पता लगा सकते हैं कि विचरण के इस उद्गम के कारण वर्ग-योग का परिमाण क्या है तो यह परिमाण अंतर-ब्लॉक वर्ग-योग में ही मिला रह जाता है और हम कहते हैं कि यह प्रभाव ब्लॉक के साथ समाकुलित (confounded) है।

§ २३.२ परस्पर-क्रिया का समाकुलन

जिस बहु-उपादानीय प्रयोग का हम पहले विवरण दे चुके हैं उसमें यदि यह पाया जाय कि एक ही ब्लॉक में आठ प्लॉट रखना उचित नहीं है तो कुल उपचार सच्यों को दो भागों में विभाजित करके चार-चार प्लॉटों के ब्लॉक बनाये जा सकते हैं। हमारे पिछले प्रयोग के हर एक ब्लॉक को दो भागों a तथा b में विभाजित किया जा सकता है। इस प्रकार प्रारम्भिक ब्लॉक को अब हम ब्लॉक-युग्म कह सकते हैं। इन कुल उपचार सच्यों को इस प्रकार विभाजित करना चाहिए कि त्रि-उपादानीय परस्पर क्रिया VSM को छोड़कर अन्य सभी मुख्य प्रभावों और परस्पर क्रियाओं का प्राक्कलन किया जा सके तथा उनके शून्य होने की निराकरणीय परिकल्पना की जाच की जा सके। इसके लिए हम उपचार-सच्यों को निम्नलिखित रूप से विभाजित कर सकते हैं।

$$a \quad \left[\begin{array}{cccc} \underline{V_1 S_1 M_1} & \underline{V_1 S_2 M_2} & \underline{V_2 S_2 M_1} & \underline{V_2 S_1 M_2} \end{array} \right]$$

$$b \quad \left[\begin{array}{cccc} \underline{V_2 S_1 M_1} & \underline{V_1 S_1 M_2} & \underline{V_1 S_2 M_1} & \underline{V_1 S_2 M_2} \end{array} \right]$$

हम यह जानते हैं कि ब्लॉक-युग्म के भाग b के उपचार सच्यों के प्रभावों के योग में से भाग a के उपचार सच्यों के प्रभावों के योग को घटाने से VSM का प्राक्कलन होता है (समी० २२७)। परन्तु क्योंकि a और b की पैदावारों में इन उपादानों के प्रभाव के अतिरिक्त ब्लॉक के प्रभाव भी शामिल हैं, इसलिए b की पैदावार में से a की पैदावार को घटाने से हमें $VSM+4(B_b-B_a)$ का अनुमान लगता है। यहाँ B_b और B_a द्वारा हम ब्लॉक b और ब्लॉक a के प्रभावों को सूचित कर रहे हैं। इस प्रकार हम देखते हैं कि त्रि-उपादानीय परस्पर क्रिया ब्लॉक प्रभावों के साथ समाकुलित है और एक ब्लॉक-युग्म द्वारा उसका अलग से अनुमान नहीं लगाया जा सकता।

अब यह देखना है कि कहीं अन्य मुख्य प्रभाव अथवा द्वि-उपादानीय परस्पर क्रियाएँ भी तो ब्लॉक प्रभावों के साथ समाकुलित नहीं हैं। इसके लिए हम एक मुख्य प्रभाव और एक द्वि-उपादानीय परस्पर-क्रिया का प्राक्कलन करने की चेष्टा करेंगे।

$$4V = (V_2 S_1 M_2 + V_2 S_2 M_1) - (V_1 S_1 M_1 + V_1 S_2 M_2) \\ + (V_2 S_1 M_1 + V_2 S_2 M_2) - (V_1 S_1 M_2 + V_1 S_2 M_1) \dots (23 I)$$

यह देखा जा सकता है कि इस परिकलन में हर एक ब्लॉक में दो प्लॉटों की पैदावार के योग में से अन्य दो प्लॉटों की पैदावार को घटाया जाता है। अतः यद्यपि प्रत्येक सच्य में ब्लॉक प्रभाव B_b या B_a भी विद्यमान है तथापि इस प्रकार के योग और वियोग से ये ब्लॉक प्रभाव हट जाते हैं और हमें मुख्य प्रभाव V का शुद्ध अनुमान

प्राप्त हो जाता है। (देखो समी० 22 I)। इसी प्रकार आप देख सकते हैं कि अन्य मुख्य प्रभावो के भी शुद्ध अनुमान प्राप्त करना संभव है।

अब हम एक द्वि-उपादान परस्पर-क्रिया का प्राक्कलन करने की चेष्टा करेंगे।

$$VS = (V_2S_2M_1 + V_1S_1M_1) - (V_1S_2M_2 + V_2S_1M_2) \\ + (V_2S_2M_2 + V_1S_1M_2) - (V_1S_2M_1 + V_2S_1M_1) \dots\dots(23 2)$$

इसमें भी ब्लॉक प्रभाव जितनी बार जोड़े जाते हैं उतनी ही बार घटा दिये जाते हैं। इस प्रकार VS के प्राक्कलन से ब्लॉक प्रभाव हट जाता है और हमें इस परस्पर क्रिया का शुद्ध प्राक्कलन बिना किसी समाकुलन (confounding) के पता चल जाता है (देखो समी० 22 4)।

§ 23.3 विदलेपण

आइये, अब हम देखें कि इस प्रयोग-अभिकल्पना में विश्लेषण किस प्रकार किया जाय। इस विश्लेषण के विभिन्न चरण निम्नलिखित हैं।

(१) कुल ब्लॉकों के लिए अंतर-ब्लॉक वर्ग-योग का परिकलन।

(२) जो मुख्य प्रभाव या परस्पर-क्रियाएँ समाकुलित नहीं हुई हैं उनके वर्गों के योग का परिकलन। यदि पहले समाकुलन का विचार किये बिना उपचार वर्ग-योग का परिकलन कर लिया गया हो तो इसमें से समाकुलित परस्पर-क्रिया के वर्ग-योग को घटाने से भी हमें यही मान प्राप्त होगा।

(३) त्रुटि वर्ग-योग को कुल वर्ग-योग में से अंतर-ब्लॉक वर्ग-योग तथा उपचार वर्ग-योग के योग को घटा कर प्राप्त करना।

निम्नले अध्याय के उदाहरण के लिए ये चरण नीचे दिये हुए हैं

(देखिए सारणी सख्या 22.I)

सारणी सख्या 23 1

VSM के समाकुलित होने पर ब्लॉक-योग

ब्लॉक	I _a	I _b	II _a	II _b
	3+7	5+6	5+10+8	6+8
योग	+6+10 =26	+5+14 =30	+12 =35	+7+15 =36
ब्लॉक	III _a	III _b	IV _a	IV _b
	4+8	5+5	4+7	4+5
योग	+6+10 =28	+7+11 =28	+4+8 =23	+5+8 =23

सारणी सख्या 23.2

VSM के समाकुलित होने पर प्रसरण विश्लेषण

विवरण का उद्गम	स्वातन्त्र्य सख्या	वर्ग-योग	वर्ग-माध्य	अनुपात	5% स्तर पर अर्थपूर्ण मान
I	II	3	4	5	6
ब्लॉक युग्म	3	$S_1=42.75$	$M_1=14.25$		
VSM	1	$S_2=0.50$	$M_2=0.50$	$\frac{M_2}{M_1} = \frac{0.50}{0.58} = 0.86$	10.13
(VSM के लिए) बूटि	3	$S_3=S_0-S_2$ $=1.75$	$M_3=0.58$		
कुल ब्लॉक	7	$S_0=45.00$	$M_0=6.43$	$\frac{M_0}{M_3} = \frac{6.43}{1.19} = 5.40$	2.58
(VSM को छोड़ कर) उपचार	6	$S_4=203.00$	$M_4=33.83$	$\frac{M_4}{M_3} = \frac{33.83}{19.1} = 28.43$	2.66
बूटि	18	$S_5=S-S_0$ $=21.50$	$M_5'=1.19$		
कुल	31	$S=269.50$			

ऊपर की सारणी में ब्लॉकयुग्म वर्ग-योग वही है जो सारणी सख्या 22.2 में ब्लॉक वर्ग-योग था क्योंकि सारणी सख्या 23.1 में ब्लॉक युग्म वही है जो सारणी सख्या 22.1 में ब्लॉक थे। उपचार वर्ग-योग दो अलग-अलग रीतियों से निकाला जा सकता है। एक तो VSM को समाकुलित न मान कर किये हुए विश्लेषण (देखो सारणी 22.2, 22.3) में प्राप्त उपचार वर्ग-योग में से VSM वर्ग-योग $\frac{1^2}{2} = 0.50$ को घटाकर।

$$203.50 - 0.50 = 203.00$$

दूसरे, जितने a ब्लॉक हैं—यानी I_a, II_a, III_a और IV_a उनमें केवल चार उपचारों के प्रयोग हैं। इसलिए इन उपचारों के अंतरों के कारण हमें एक उपचार

वर्ग-योग प्राप्त हो सकता है जिसकी स्वातन्त्र्य सख्या 3 है। इसी प्रकार b ब्लॉकों में से हम अन्य उपचारों के अंतर से प्राप्त वर्ग योग का परिकलन कर सकते हैं जिसकी स्वातन्त्र्य सख्या भी 3 है। इन दोनों के योग से हमें ब्लॉक के अंतर का कुल उपचार वर्ग-योग प्राप्त होता है जिसकी स्वातन्त्र्य सख्या 6 है। सारणी 22 I के अनुसार Σ ब्लॉकों के 16 प्लॉटों की कुल पैदावार 112 तथा a ब्लॉकों के लिए उपचार वर्ग-योग

$$S_{aa} = [(16 \times 4) + (32 \times 8) + (24 \times 6) + (40 + 10)] - \frac{(112)^2}{16}$$

$$= 864 - 784$$

$$= 80$$

b ब्लॉकों के 16 प्लॉटों की कुल पैदावार = 116 तथा b ब्लॉकों के लिए उपचार वर्ग-योग $S_{bb} = [(20 \times 5) + (24 \times 6) + (24 \times 6) + (48 \times 12)] - \frac{(116)^2}{16}$

$$= 964 - 841$$

$$= 123$$

$$\text{इस प्रकार कुल उपचार वर्ग-योग} = 80 + 123$$

$$= 203$$

वास्तव में a ब्लॉकों और b ब्लॉकों के लिए अलग-अलग विश्लेषण किया जा सकता है। इसके द्वारा दोनों उपचार वर्ग योगों को जोड़ कर कुल उपचार वर्ग-योग, तथा त्रुटि वर्ग योगों को जोड़ कर कुल त्रुटि-वर्ग योग प्राप्त किया जा सकता है। ब्लॉक वर्ग-योग के लिए हमें एक पद और जोड़ना चाहिए जो a ब्लॉकों और b ब्लॉकों के बीच के अंतर से संबंधित है।

a ब्लॉकों के लिए विश्लेषण

$$(1) \text{ ब्लॉक वर्ग योग } S_{1a} = \frac{(26)^2 + (35)^2 + (28)^2 + (23)^2}{4} - \frac{(112)^2}{16}$$

$$(\text{देखिए सारणी सख्या 23 I}) = \frac{676 + 1225 + 764 + 929}{4} - 784$$

$$= 803.5 - 784$$

$$= 19.5$$

$$(ii) \text{ कुल वर्ग योग } S_a = [3^2 + 5^2 + 4^2 + 4^2 + 7^2 + 10^2 + 8^2 + 7^2 \\ (\text{देखिए सारणी सख्या 22 I}) + 6^2 + 8^2 + 6^2 + 4^2 + 10^2 + 12^2 + 10^2 + 8^2] \\ - \frac{(112)^2}{16}$$

$$= 888 - 784$$

$$= 104$$

b ब्लॉक के लिए विप्लेषण

$$(i) \text{ ब्लॉक वर्ग-योग } S_{1b} = \frac{(30)^2 + (36)^2 + (28)^2 + (22)^2}{4} - \frac{(116)^2}{16} \\ (\text{देखिए सारणी सख्या 23.1})$$

$$= \frac{3464}{4} - 841$$

$$= 866 - 841$$

$$= 25$$

$$(ii) \text{ कुल वर्ग योग } S_b = [5^2 + 6^2 + 5^2 + 4^2 + 6^2 + 8^2 + 5^2 + 5^2 \\ (\text{देखिए सारणी सख्या 22 I}) + 5^2 + 7^2 + 5^2 + 14^2 + 14^2 + 15^2 + 11^2 + 8^2] \\ - \frac{(116)^2}{112}$$

$$= 1006 - 841$$

$$= 165$$

इस सारणी (सारणी अगले पृष्ठ पर देखिए) सख्या 23.3 में ब्लॉक-वर्ग योग तथा कुल-वर्ग-योग के लिए अतिथि स्तम्भ में a और b ब्लॉकों में विभाजन से उत्पन्न पद 0.5 को जोड़ने से हमें पूर्व कलन सारणी प्राप्त होती है।

ब्लॉक वर्ग-योग को दो प्रकार से विभाजित किया जा सकता है जैसा ऊपर की दो सारणियों द्वारा स्पष्ट है। पहली सारणी में विभाजन यह समझ कर किया जा सकता है कि ब्लॉक-युग्म तो ब्लॉक है और उसके दो भाग प्लॉट। इस प्रकार कुल ब्लॉक वर्ग-योग को अंतर ब्लॉक युग्म, द्रुति तथा उपचार वर्ग-योग में बाँटा जा सकता है। यह उपचार वर्ग-योग VSM के कारण है। इस प्रकार के विभाजन से VSM के वर्ग-योग को भी जाँचा जा सकता है, परंतु इसके लिए द्रुति अंतर-ब्लॉक-युग्म वर्ग-योग से

सारणी सारथा 23.3

प्रसरण वितरलेखन सारणी

विचरण का उद्गम	a ब्लॉक			b ब्लॉक			कुल
	स्वातंत्र्य संख्या	वग योग	स्वातंत्र्य संख्या	वग योग	स्वातंत्र्य संख्या	वग योग	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	
ब्लॉक	3	$S_{2a} = 195$	3	$S_{1b} = 250$	6	$S_b - S_2 = S_{1a} + S_{1b} = 445$	
उपचार	3	$S_{2b} = 800$	3	$S_{1a} = 1230$	6	$S_2 = S_{2a} + S_{2b} = 2030$	
दुटि	9	$S_{ra} = S_a - S_{1a} - S_{2a} = 45$	9	$S_{rb} = S_b - S_{1b} - S_{2b} = 170$	18	$S_e = S_{ra} + S_{rb} = 215$	
कुल	15	$S_a = 1040$	15	$S_b = 1650$	30	$S = S_a + S_b = 2690$	

प्राप्त होती है। दूसरी सारणी में विभाजन अंतर-*a* ब्लॉक, अंतर-*b* ब्लॉक तथा *a* और *b* ब्लॉकों के माप्यों के अंतर द्वारा किया गया है।

ऊपर के कुछ पृष्ठों से आपको यह मालूम हुआ होगा कि यद्यपि एक ही प्रयोग द्वारा समाकुलित परस्पर क्रिया का प्राक्कलन संभव नहीं है, परंतु कई बार किये हुए प्रयोगों द्वारा यह संभव है। इस समाकुलित परस्परक्रिया के प्राक्कलन की त्रुटि अन्य प्राक्कलनों की त्रुटि से अधिक होती है और इस त्रुटि की स्वातंत्र्य सध्या भी बहुत कम रह जाती है। ऊपर हमने इस प्रकार की अभिकल्पना का वर्णन किया है जिसमें केवल एक परस्पर क्रिया *VSM* प्रत्येक ब्लॉक युग्म में समाकुलित है। इसके अतिरिक्त ऐसी अभिकल्पना भी की जा सकती है जिसमें समाकुलन संपूर्ण न होकर केवल आंशिक हो।

§ २३.४ आंशिक समाकुलन (*Partial confounding*)

इस प्रकार की अभिकल्पना में भिन्न-भिन्न ब्लॉक-युग्मों में भिन्न-भिन्न परस्पर क्रियाओं को ब्लॉक-प्रभावों से समाकुलित किया जाता है। इस प्रकार यदि एक परस्पर क्रिया एक ब्लॉक युग्म में ब्लॉक-प्रभावों से समाकुलित है तो उसका प्राक्कलन दूसरे ब्लॉक युग्मों द्वारा लगाया जा सकता है। इस प्रकार की अभिकल्पना का एक उदाहरण नीचे दिया हुआ है।

सारणी संख्या 23.4

आंशिक समाकुलित अभिकल्पना-उपचारों का अनुक्रम और ब्लॉक-योग

समाकुलित परस्पर क्रिया	<i>VSM</i>		<i>VM</i>		<i>VS</i>		<i>MS</i>	
	<i>I_a</i>	<i>I_b</i>	<i>II_a</i>	<i>II_b</i>	<i>III_a</i>	<i>III_b</i>	<i>IV_a</i>	<i>IV_b</i>
ब्लॉक	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1)	$V_1 M_1 S_1$	$V_2 M_1 S_1$	$V_2 M_1 S_1$	$V_1 M_1 S_1$	$V_2 M_1 S_1$	$V_1 M_1 S_1$	$V_2 M_2 S_1$	$V_1 M_1 S_1$
	$V_2 M_2 S_1$	$V_1 M_2 S_1$	$V_1 M_2 S_1$	$V_2 M_1 S_2$	$V_1 M_2 S_2$	$V_1 M_2 S_1$	$V_1 M_1 S_2$	$V_2 M_1 S_1$
	$V_2 M_1 S_2$	$V_1 M_1 S_2$	$V_2 M_1 S_2$	$V_2 M_2 S_1$	$V_1 M_2 S_1$	$V_2 M_1 S_2$	$V_2 M_2 S_1$	$V_1 M_2 S_2$
	$V_1 M_2 S_2$	$V_2 M_2 S_2$	$V_1 M_2 S_2$	$V_2 M_2 S_2$	$V_1 M_2 S_2$	$V_2 M_2 S_2$	$V_2 M_1 S_1$	$V_2 M_2 S_2$
ब्लॉक योग	26	30	36	35	26	30	21	24

§ २३.५ सांख्यिकीय विदलेपण

आंशिक समाकुलन की स्थिति में जिस साधारण नियम का पालन किया जाता है वह केवल यह है कि आंशिक समाकुलित परस्पर-क्रियाओं का प्राक्कलन उन ब्लॉक-युग्मों से लगाया जाता है जिनमें वे समाकुलित नहीं हैं। इन प्राक्कलनों से वर्ग-योग उसी प्रकार परिकलित किया जाता है जैसे अनसमाकुलित अभिकल्पनाओं में। यह ध्यान में रखना होता है कि ये अनुमान कम प्लॉटों पर आधारित हैं। ब्लॉक वर्ग-योग का परिकलन ब्लॉक योगों के आधार पर साधारण तरीके से ही किया जाता है।

यदि हमने परस्पर-क्रियाओं के योग का परिकलन—बिना समाकुलन का ध्यान रखते हुए ही सब ब्लॉक-युग्मों के आधार पर कर लिया हो तो इस परिकलित मान में से उन ब्लॉक-युग्मों का अंतर घटा कर इसे ठीक किया जा सकता है जिनमें वे समाकुलित हैं। ऊपर के उदाहरण में यदि परस्पर-क्रिया V/M के योग का परिकलन करता है तो वह पुराने योग में ब्लॉक II_2 के योग को जोड़ कर तथा II_3 के योग को घटा कर किया जा सकता है।

इस प्रकार

$$[VM]' = -4 + 36 - 35 = -3$$

$$[VS]' = 13 + 26 - 30 = 8$$

$$[MS]' = 20 + 21 - 24 = 17$$

$$[VSM]' = 4 + 26 - 30 = 0$$

प्रसरण विदलेपण में अब हर एक परस्पर-क्रिया के लिए एक एक स्वातंत्र्य-संख्या होगी क्योंकि इन सबका प्राक्कलन किया जा सकता है। परस्पर-क्रियाओं के वर्ग-योग ऊपर दिये हुए योगों के वर्ग को 24 से विभाजित करने से मिलते हैं क्योंकि इनमें से प्रत्येक 24 प्लॉटों की उपजों के योग और वियोग द्वारा परिकलित है। जिस जिस ब्लॉक-युग्म में वे समाकुलित हैं उनके आठ प्लॉटों का उपयोग इनके परिकलन में नहीं किया गया है। मुख्य प्रभावों का वर्ग-योग वही रहता है जो पहले था। ब्लॉक वर्ग-योग का कलन ब्लॉक योगों से किया जाता है और अब में त्रुटि वर्ग-योग को घटाकर मालूम कर लिया जाता है।

$$VM \text{ के कारण वर्ग योग} = \frac{3^2}{24} = 0.375$$

$$VS \text{ के कारण वर्ग योग} = \frac{8^2}{24} = 2.667$$

$$MS \text{ के कारण वर्ग-योग} = \frac{(17)^2}{24} = 12\ 042$$

$$VSM \text{ के कारण वर्ग-योग} = \frac{0^2}{24} = 0\ 000$$

$$V \text{ के कारण वर्ग-योग} = \frac{(5 \times 4)^2}{32} = 12\ 500 \quad (\text{देखिए सारणी सख्या २२३})$$

$$S \text{ के कारण वर्ग-योग} = \frac{(15 \times 4)^2}{32} = 112\ 500$$

$$M \text{ के कारण वर्ग-योग} = \frac{(11 \times 4)^2}{32} = 60\ 500$$

$$\begin{aligned} \text{ब्लॉक वर्ग-योग} &= \frac{1}{4} [(26)^2 + (30)^2 + (36)^2 + (35)^2 + (26)^2 + (30)^2 \\ &\quad + (21)^2 + (24)^2] - \frac{(228)^2}{16} \\ &= 48\ 000 \end{aligned}$$

सारणी सख्या 23 5

भातक-समाकुलित अभिकल्पना का प्रसरण विश्लेषण

विचरण का उद्गम	स्वातंत्र्य सख्या	वर्ग-योग	वर्ग-माध्य	अनुपात	5% स्तर पर अनुपात का बर्धपूर्ण मान
(1)	(2)	(3)	(4)	(5)	(6)
ब्लॉक	7	48 000	6 857	5 575	2 62
V	1	12 500	12 500	10 163	4 45
M	1	60 500	60 500	49 187	4 45
S	1	112 500	112 500	91 464	4 45
मुख्य प्रभाव	3	185 500	61 833	50 271	3 20
VM	1	0 375	0 375	0 305	4 45
VS	1	2 667	2 667	2 168	4 45
MS	1	12 042	12 042	9 790	4 45
VSM	1	0 000	0 000	0 000	4 45
परस्पर क्रिया	4	15 084	3 771	3 060	2 96
त्रुटि	17	20 916	1 230		
कुल	31	269 500			

अध्याय २४

संतुलित असंपूर्ण ब्लॉक अभिकल्पना

Balanced Incomplete Block Design

§ २४.१ परिभाषा

पिछले अध्याय में हमने कुछ असंपूर्ण ब्लॉक अभिकल्पनाओं से परिचय प्राप्त किया था जिनका प्रयोग बहु-उपादानीय प्रयोगों में किया जाता है। इस अध्याय में हम एक अन्य प्रकार की असंपूर्ण-ब्लॉक अभिकल्पना का अध्ययन करेंगे जिसको संतुलित असंपूर्ण ब्लॉक अभिकल्पना कहा जाता है। इस अभिकल्पना के कुछ नियम हैं जो नीचे दिये हुए हैं।

(1) हर एक ब्लॉक में प्लॉटों की संख्या बराबर होती है। इस संख्या को हम k से सूचित करेंगे।

(2) हर एक उपचार का जितने ब्लॉकों में पुनः प्रयोग किया जाय उतनी संख्या बराबर होती है। इस पुनः प्रयोग की संख्या को हम r से सूचित करेंगे। एक ब्लॉक में एक उपचार का एक ही बार प्रयोग होता है।

(3) उपचारों में से यदि दो-दो के युग्म बनाये जायें तो हर एक युग्म के उपचार किमी न किमी ब्लॉक में अवश्य साय-साय आते हैं। उन ब्लॉकों की संख्या जिनमें किमी विभेद युग्म के उपचार साय-साय आते हैं प्रत्येक युग्म के लिए समान होती है। इस संख्या को हम λ से सूचित करेंगे।

कुल उपचारों की संख्या को हम v से और कुल ब्लॉकों की संख्या को b से सूचित करेंगे। इसके पहले कि हम इस प्रकार की अभिकल्पना का उदाहरण सहित विस्तरेण करें, हमको अधिक स्पष्ट करने के लिए एक-दो सरल उदाहरण नीचे दिये जाते हैं।

§ २४.२ उदाहरण

ऊपर दिये हुए नियमों से, विभेदकर तीसरे नियम से, स्पष्ट है कि एक ब्लॉक में कम से कम दो प्लॉट अवश्य होने चाहिए। यदि कुल उपचार पाँच हों जिन्हें A, B, C, D और E से सूचित किया जाय तो तीसरे नियम के अनुसार प्रत्येक उपचार-युग्म कम-से-कम एक ब्लॉक में अवश्य होना चाहिए।

1. यदि एक ब्लॉक में केवल दो प्लॉट हो तो अभिकल्पना में कम से कम दस प्लॉट अवश्य होने चाहिए जिनमें (1) AB (2) AC (3) AD (4) AE (5) BC (6) BD (7) BE (8) CD (9) CE तथा (10) DE ये दस उपचार-गुम होंगे। या हो सकता है कि प्रत्येक समूह को दो या तीन बार दुहराया गया हो। कुल भी हो, यदि कुल उपचारों की संख्या पांच है और हर एक ब्लॉक में केवल दो प्लॉट हैं तो कुल ब्लॉकों की संख्या $(\frac{v}{2}) = 10$ अथवा दस का कोई गुणज (multiple) होगी।

ii उपर्युक्त स्थिति एक सीमान्त स्थिति है क्योंकि दो से कम प्लॉट किसी संतुलित असंपूर्ण अभिकल्पना में हो ही नहीं सकते। दूसरी सीमान्त स्थिति वह होगी जब एक ब्लॉक में प्लॉटों की संख्या k कुल उपचारों की संख्या v से केवल एक कम हो। $k = v - 1$

ऊपर के पाँच उपचारों में से चार-चार एक-एक ब्लॉक में हो और तीसरा नियमों का पालन हो तो यह दसका एक उदाहरण होगा। इस स्थिति में कुल ब्लॉकों की संख्या b पाँच या पाँच का कोई गुणज होगी। ये चार-चार के पाँच समूह निम्नलिखित हैं : (1) $ABCD$ (2) $ABCE$ (3) $ABDE$ (4) $ACDE$ (5) $BCDE$

क्योंकि प्रत्येक ब्लॉक में एक उपचार का प्रयोग नहीं होता और क्योंकि प्रत्येक उपचार का पुनः प्रयोग समान संख्या में होना चाहिए, इसलिए यह स्पष्ट है कि इन पाँचों संयोजनों (combinations) का बराबर संख्या में होना संतुलित असंपूर्ण ब्लॉक अभिकल्पना के लिए आवश्यक है।

ऊपर की अभिकल्पना में

$$k=4, r=4, \lambda=3, v=5, b=5$$

आपको यह भ्रम हो सकता है कि यदि एक ब्लॉक में प्लॉटों की संख्या k है और कुल उपचारों की संख्या v है तो ब्लॉकों की संख्या $b = \binom{v}{k}$ होना चाहिए। ऊपर

के दोनों उदाहरणों में ऐसा हुआ था, परंतु वे दोनों सीमांत स्थितियाँ थीं। $\binom{v}{k}$

ब्लॉकों का होना उसी अवस्था में आवश्यक है जब k परिमाण का प्रत्येक सचय किसी न किसी ब्लॉक में अवश्य हो। किन्तु असंपूर्ण ब्लॉक अभिकल्पना में अनेक सचय किसी भी ब्लॉक में नहीं होते।

3 मान लीजिए, कुल उपचारों की संख्या सात है और एक एक ब्लॉक में तीन तीन प्लॉट हैं। नीचे एक अभिकल्पना दी जाती है। यह देखना है कि यह एक संतुलित असंपूर्ण अभिकल्पना है या नहीं।

$ABD, ACE, CDG, AGF, BCF, BEG, DEF$

(1) क्योंकि प्रत्येक ब्लॉक में प्लॉटों की संख्या तीन है इसलिए पहिले नियम का पालन हो रहा है।

(2) हर एक उपचार का पुन प्रयोग तीन तीन बार हो रहा है इसलिए दूसरे नियम का पालन हो रहा है।

(3) दो दो के जो इक्कीस समूह इन सात उपचारों से बनाये जा सकते हैं वे सब किसी न किसी ब्लॉक में अवश्य पाये जाते हैं और एक उपचार-युग्म एक से अधिक ब्लॉकों में भी नहीं पाया जाता। आप यह देख सकते हैं कि किन्हीं भी दो ब्लॉकों में दो उपचार एक-से नहीं हैं। इस प्रकार तीसरे नियम का भी पालन हो रहा है। इसलिए परिभाषा के अनुसार यह अभिकल्पना एक संतुलित असंपूर्ण ब्लॉक अभिकल्पना है।

इस अभिकल्पना में ब्लॉकों की संख्या केवल 7 है, न कि $\binom{7}{2} = 35$ ।

§ २४.३ संतुलित असंपूर्ण ब्लॉक अभिकल्पना के प्राचलो के कुछ संबंध

किसी भी संतुलित असंपूर्ण-अभिकल्पना को b, k, r, v और λ द्वारा सूचित किया जा सकता है जो इसके प्राचल हैं। आप इन संकेतों से पहले से ही परिचित हैं।

क्योंकि कुल ब्लॉकों की संख्या b है और प्रत्येक ब्लॉक में k प्लॉट है इसलिए कुल प्लॉटों की संख्या bk है।

क्योंकि कुल उपचारों की संख्या v है और हर एक उपचार का r प्लॉटों में पुनः प्रयोग किया गया है इस कारण कुल प्लॉटों की संख्या को vr द्वारा भी सूचित किया जा सकता है।

$$\therefore bk = vr \quad (A)$$

इसके अतिरिक्त जिन ब्लॉकों में कोई एक विशेष उपचार (यथा A) मौजूद हो उनकी संख्या है r , और इस प्रकार के प्रत्येक ब्लॉक में $k-1$ ऐसे प्लॉट हैं जिनमें यह विशेष उपचार मौजूद नहीं है। अतः इन ब्लॉकों में जिन प्लॉटों में A मौजूद नहीं है उनकी संख्या होगी $r(k-1)$ — परंतु यही वे ब्लॉक हैं जिनमें इस उपचार विशेष A के साथ अन्य उपचारों के युग्म पाये जा सकते हैं। क्योंकि कुल $(v-1)$ अन्य उपचार हैं और उनमें से प्रत्येक के साथ A के λ उपचार युग्म बनते हैं, इसलिए इन्हीं

ब्लॉकों के उन प्लॉटों की संख्या जिनमें यह विशेष उपचार नहीं है $\lambda (\nu - 1)$ भी होगी।

$$\text{अतः } \lambda (\nu - 1) = r (k - 1)$$

$$\text{अथवा } \lambda = \frac{r (k - 1)}{(\nu - 1)} \dots \dots (B)$$

इसलिए संतुलित असंपूर्ण ब्लॉक अभिकल्पना के लिए $\frac{r (k - 1)}{\nu - 1}$ पूर्ण संख्या

(Integral number) होनी चाहिए। यदि हम देखें कि कोई अभिकल्पना उपयुक्त दोनों गतों A और B को पूरा करती है तो हम समझ सकते हैं कि वह संतुलित असंपूर्ण ब्लॉक अभिकल्पना है।

§ २४.४ यादुच्छिकीकरण

किसी प्रयोग के लिए उपचारों के सचयों को यादुच्छिकीकरण द्वारा विभिन्न ब्लॉकों में वितरित करना और एक सचय के उपचारों को ब्लॉक के विभिन्न प्लॉटों में यादुच्छिकीकरण द्वारा वितरित करना आवश्यक है।

§ २४.५ खेती से संबंधित एक संतुलित-असंपूर्ण ब्लॉक अभिकल्पना

आइए, अब हम देखें कि एक संतुलित असंपूर्ण ब्लॉक अभिकल्पना का विश्लेषण किस प्रकार किया जाता है। दूसरी अभिकल्पनाओं की भाँति इसको भी उदाहरण द्वारा समझाया जायगा।

§ २४.५.१ विश्लेषण के लिए प्रतिरूप, प्रतिरूप के प्राचलों वग प्राचकलन

यह देखने के लिए कि उनकी पैदावारों में कुछ विशेष अंतर है अथवा नहीं, पाँच प्रकार के गेहूँ के बीजों पर प्रयोग किया जा रहा है। यदि ब्लॉक i में किरम j के गेहूँ की पैदावार को y_{ij} से सूचित किया जाय तो प्रतिरूप के अनुसार

$$E (y_{ij}) = h_i + t_j \dots \dots (24.1)$$

$$\text{और } \sum_{j=A}^E t_j = 0 \dots \dots (24.2)$$

यहाँ b_j द्वारा i वें ब्लॉक के प्रभाव और t_j द्वारा j वीं किस्म के प्रभाव को सूचित किया जा रहा है। j वीं किस्म के प्रभाव t_j से हमारा तात्पर्य j वीं किस्म के गेहूँ की पैदावार तथा सब किस्मों की औसत पैदावार के अंतर के प्रत्याशित मान से है। इसी कारण हमें समीकरण (24.2) प्राप्त होता है। मान लीजिए अभिकल्पना में पाँच ब्लॉक हैं जिनमें निम्नलिखित उपचार समूह हैं

$$(1) ABCD \quad (2) ABCE \quad (3) ABDE \quad (4) ACDE \quad (5) BCDE$$

यदि i -वें ब्लॉक की कुल पैदावार को II_i से सूचित किया जाय तो

$$\left. \begin{aligned} E(B_1) &= 4b_1 + t_A + t_B + t_C + t_D \\ E(B_2) &= 4b_2 + t_A + t_B + t_C + t_E \\ E(B_3) &= 4b_3 + t_A + t_B + t_D + t_E \\ E(B_4) &= 4b_4 + t_A + t_C + t_D + t_E \\ E(B_5) &= 4b_5 + t_B + t_C + t_D + t_E \end{aligned} \right\} \quad (C)$$

यहाँ $4 = k$ प्रत्येक ब्लॉक के प्लॉटों की संख्या है।

इसके अतिरिक्त यदि T_j द्वारा उन प्लॉटों की पैदावार के योग को सूचित किया जाय जिसमें j -वीं किस्म बोयी गयी है तो

$$\left. \begin{aligned} E(T_A) &= 4t_A + b_1 + b_2 + b_3 + b_4 \\ E(T_B) &= 4t_B + b_1 + b_2 + b_3 + b_5 \\ E(T_C) &= 4t_C + b_1 + b_2 + b_4 + b_5 \\ E(T_D) &= 4t_D + b_1 + b_3 + b_4 + b_5 \\ E(T_E) &= 4t_E + b_2 + b_3 + b_4 + b_5 \end{aligned} \right\} \quad (D)$$

यहाँ $4 = r =$ प्रत्येक किस्म के पुनः प्रयोग की संख्या है।

$$\therefore E \left[T_A - \frac{B_1 + B_2 + B_3 + B_4}{4} \right]$$

$$= 4t_A + b_1 + b_2 + b_3 + b_4 - \frac{4(b_1 + b_2 + b_3 + b_4) + t_A + 3 \sum_{j=A}^E t_j}{4}$$

$$= \frac{15}{4} t_A - \frac{3}{4} \sum_{j=A}^E t_j$$

परन्तु क्योंकि $\sum_{j=A}^E t_j = 0$ इसलिए

$$E \left[T_A - \frac{B_1 + B_2 + B_3 + B_4}{4} \right] = \frac{15}{4} t_A$$

इसलिए यदि $T_A - \frac{B_1 + B_2 + B_3 + B_4}{4}$ को Q_A से सूचित किया जाय तो

$$t_A \text{ का प्राक्कलक } \hat{t}_A = \frac{4}{15} Q_A \text{ है।}$$

इस उदाहरण की अभिकल्पना में

$$k=4, b=5, v=5, r=4, \lambda=3$$

$$\therefore \hat{t}_A = \frac{k}{\lambda v} Q_A$$

$$\text{इसी प्रकार } \hat{t}_j = \frac{k}{\lambda v} Q_j \quad j=A, B, C, D, E, \dots (E)$$

जहाँ $Q_j = T_j -$ (उन ब्लॉकों की औसत पैदावार जिनमें j -वी किस्म बोयी गयी है)।

यह अधिक साधारण मूल है और हम प्रकार की किसी भी अभिकल्पना में इसका उपयोग हो सकता है।

Q_j को समजित उपचार योग (adjusted treatment total) कहा जाता है क्योंकि इसमें ब्लॉकों का प्रभाव हटा दिया जाता है।

§ २४.५.२ परिकल्पना परीक्षण

इस \hat{t}_j के प्रसरण को हम $\frac{k}{\lambda v} \sigma^2$ से सूचित करेंगे। क्योंकि \hat{t}_j और \hat{t}_j स्वतंत्र हैं इसलिए

$$V(\hat{t}_j - \hat{t}_{j'}) = \frac{2k}{\lambda v} \sigma^2 \quad \dots (24.3)$$

हम t -परीक्षण द्वारा t_j और $t_{j'}$ के अंतर से संबंधित परिकल्पनाओं की जाँच कर सकते हैं। परंतु इसके लिए σ^2 के अनुमान का ज्ञात होना आवश्यक है। इसके लिए प्रसरण विश्लेषण सारणी की सहायता लेनी पडती है।

सारणी संख्या 24.1

संतुलित असंतूर्ण ब्लॉक अभिकल्पना के लिए प्रसरण विह्लेयक सारणी

विचरण का उद्गम	स्वातन्त्र्य मर्यादा	वर्ग-योग
(1)	(2)	(3)
उपचारों का उपस्था करके ब्लॉक	$b-1$	$\frac{1}{k} \sum_{i=1}^b B_i^2 - \frac{G^2}{bk}$
ब्लॉकों का प्रभाव हटाकर उपचार	$v-1$	$\sum_{j=1}^v t_j Q_j$ $= \frac{k}{\lambda v} \sum_{j=1}^v Q_j^2$
शुद्धि	$(bk-1) - [(b-1) - (v-1)]$ $= bk - b - v + 1$	• •
कुल	$bk-1$	$\sum_{i=1}^b \sum_{j=1}^v y_{ij}^2 - \frac{G^2}{bk}$

शुद्धि वर्ग-योग को कुल वर्ग-योग में से अन्य दो वर्ग-योगों को घटाकर निकाला जाता

है। इन सारणी में $G = \sum_{i=1}^b \sum_{j=1}^v y_{ij}^2$ शुद्धि वर्ग-योग में उसकी स्वातन्त्र्य

मर्यादा $bk - b - v + 1$ का भाग देने से हमें σ^2 का अनुमान होता है। इसी अनुमान का परिकल्पनाओं की जाँच में प्रयोग होता है।

§ २४.५-३ आंकड़े

गइए, अब फिर अपना ध्यान उदाहरण पर लगाया जाय।

सारणी संख्या 24.2

प्रयोग का फल

ब्लॉक 1	A	B	C	D	$6B_1 = 31$
	3	10	12		
ब्लॉक 2	A	B	C	E	$4B_2 = 29$
	4	9	12		
ब्लॉक 3	A	B	D	E	$6B_3 = 30$
	7	12	5		
ब्लॉक 4	A	C	D	E	$5B_4 = 27$
	6	9	7		
ब्लॉक 5	B	C	D	E	$9B_5 = 47$
	17	11	10		
					$G = 164$

§ 24.4.4 विश्लेषण

$$Q_A = 3+4+7+6 - \frac{31+29+30+27}{4}$$

$$= -9.25$$

$$Q_B = 10+9+12+17 - \frac{31+29+30+47}{4}$$

$$= 13.75$$

$$Q_C = 12+12+9+11 - \frac{31+29+27+47}{4}$$

$$= 10.50$$

$$Q_D = 6+5+7+10 - \frac{31+30+27+47}{4}$$

$$= -5.75$$

$$Q_E = 4+6+5+9 - \frac{29+30+27+47}{4}$$

$$= -9.25$$

$$\sum_{i=1}^5 \sum_{j=A}^E y_{ij}^2 = 1582$$

$$\sum_{i=1}^5 B_i^2 = 5640$$

$$\frac{1}{4} \sum_{i=1}^5 B_i^2 = 1410$$

$$G^2 = \left[\sum_{j=1}^5 \sum_{j=A}^E y_{ij} \right]^2 = 26869 \quad \frac{G^2}{5 \times 4} = 1344.8$$

$$\sum_{j=A}^E Q_j^2 = 4179475 \quad \sum_{j=A}^E Q_j \hat{y}_j = \frac{1}{4} \sum_{j=A}^E Q_j^2 = 11145$$

सारणी संख्या 243

प्रसरण विश्लेषण सारणी

विवरण का उद्गम	स्वातंत्र्य संख्या	वग-योग	वर्ग-माध्य	अनु-पात	5% स्तर पर अर्थ-पूर्ण मान
(1)	(2)	(3)	(4)	(5)	(6)
उपचारों की उपेक्षा करके ब्लॉक	4	65.20	16.30		
ब्लॉक प्रभाव हटाकर उपचार	4	111.45	27.86	5.07	3.36
त्रुटि	11	60.55	05.50		
कुल	19	237.20			

सहकारी चर (Concomitant Variable) का उपयोग और सह-प्रसरण विश्लेषण (Analysis of Covariance)

§ २५.१ प्रयोग को अधिक दक्ष बनाने का प्रयत्न

आप यादृच्छिकीकृत ब्लॉक अभिकल्पना, लैटिन-बर्ग अभिकल्पना आदि के अध्ययन में यह समझ ही चुके हैं कि ब्लॉक बनाने का उद्देश्य त्रुटि को कम करना है। इन अभिकल्पनाओं का विरलेषण इस अभिवारणा को लेकर किया जाता है कि यदि प्रेक्षित मान में से ब्लॉक आदि के प्रभावों को हटा दिया जाय तो शेष भाग एक यादृच्छिक चर होता है जिसका माध्य शून्य और घटन प्रसामान्य माना जा सकता है। इस घटन के प्रसरण को ही त्रुटि-वर्ग माध्य कहा जाता है। यदि हम कुछ प्रभावों को नहीं हटा पाते तो उनका वर्ग-योग भी त्रुटि वर्ग-योग में मिलकर उसे बढ़ा देता है। इस प्रकार उपचारी के प्रभावों के प्राक्कलन अक्षर (inefficient) हो जाते हैं।

त्रुटि को कम करने का एक और उपाय है। मान लीजिए, आप किसी विधौष लक्षण (characteristic) y में दिलचस्पी रखते हैं। परंतु प्रयोग में y के अतिरिक्त एक अन्य लक्षण x पर भी प्रेक्षण किये जाते हैं। यदि x का y के साथ लगभग एक-पाठ संबंध (linear relation) हो तो y के प्रेक्षण में से x के प्रभाव को हटाया जा सकता है और इस प्रकार y के ऊपर उपचार के प्रभाव को अधिक दक्षता के साथ प्राक्कलित किया जा सकता है। यह संभव है कि यह लक्षण x इस प्रकार का हो कि उसके आधार पर ब्लॉक बनाना बहुत कठिन हो। इसलिए उसके प्रभाव को ब्लॉक निर्माण द्वारा नहीं बल्कि किसी और ही तरीके से हटाया जाता है।

§ २५.२ समाश्रयण प्रतिरूप

पहले x और y के बीच एक समाश्रयण रेखा (regression line) का अनुमान लगाया जा सकता है। हम इस अभिवारणा को लेकर चलते हैं कि इस रेखा से y के विचलनों का घटन प्रसामान्य है। इस प्रसामान्य घटन के प्रसरण को ही हम त्रुटि-वर्ग माध्य कहेंगे।

यदि 1-वें ब्लॉक में j -वें उपचार पानेवाले प्लॉट के लिए y लक्षण का मान y_{ij} तथा x लक्षण का मान x_{ij} हो तो इस प्रतिरूप के अनुसार

$$y_{ij} = \mu + b_j + t_i + \beta(x_{ij} - \bar{x}) + \epsilon_{ij}$$

$$i=1, 2, \quad b$$

$$j=1, 2, \quad v \quad (25.1)$$

जहाँ $\mu = Y_{ij}$ के प्रत्याक्षित मानों का माध्य

पिछले विश्लेषणों की भाँति हम यह अधिधारणा लेकर चल सकते हैं कि

$$\sum_{i=1}^v t_i = 0 \quad (25.2)$$

तथा

$$\sum_{j=1}^b b_j = 0 \quad (25.3)$$

§ २५.३ उपचारों के प्रभाव समान होने की परिकल्पना के अंतर्गत समाश्रयण प्रतिरूप के प्राचली का प्राक्कलन

यदि हमें इस निराकरणीय परिकल्पना की जाँच करनी है कि सब उपचारों के प्रभाव समान हैं तो इसके अनुसार

$$t_i = 0, \quad j=1, 2, \quad v$$

इस परिकल्पना के अन्तगत समीकरण (25.1) बतल कर निम्नलिखित हो जायगा

$$Y_{ij} = \mu + b_j + \beta(x_{ij} - \bar{x}) + \epsilon_{ij} \quad (25.4)$$

हम नीचे निम्नलिखित सकेतों का उपयोग करेंगे

$$Y_i = \sum_{j=1}^v Y_{ij}, \quad X_i = \sum_{j=1}^v x_{ij}$$

$$Y_j = \sum_{i=1}^b y_{ij}, \quad X_j = \sum_{i=1}^b x_{ij}$$

$$Y = \sum_{i=1}^b Y_i = \sum_{j=1}^v Y_j = \sum_{i=1}^b \sum_{j=1}^v Y_{ij}, \quad \bar{y} = \frac{Y}{bv}$$

$$X = \sum_{i=1}^b X_i = \sum_{j=1}^v X_j = \sum_{i=1}^b \sum_{j=1}^v x_{ij}, \quad \bar{x} = \frac{X}{bv}$$

हमें μ , b , और β का प्राक्कलन करना है जहाँ $i=1, 2, \dots, b$ । यदि इनके प्राक्कलनों को क्रमशः $\hat{\mu}$, \hat{b} , तथा $\hat{\beta}$ से सूचित किया जाय तो इनके लिए हमें निम्न लिखित समीकरण प्राप्त होते हैं।

(1) $b\hat{\mu} = Y \quad \Rightarrow \quad$ सब y -प्रेक्षणों का योग

अथवा $\hat{\mu} = \frac{Y}{b} = \bar{y} \quad (255)$

(2) $\nu(\hat{\mu} + \hat{b}_i) + \hat{\beta} \left[X_i - \frac{X}{b} \right] = Y_i$

$=$ i -वें ब्लॉक के y -प्रेक्षण का योग

अथवा

$$\hat{b}_i = \frac{1}{\nu} \left(Y_i - \frac{Y}{b} \right) - \frac{\hat{\beta}}{\nu} \left[X_i - \frac{X}{b} \right] \quad (256)$$

$i=1, 2, \dots, b$

(3) $\sum_{i=1}^b \hat{b}_i \left[X_i - \frac{X}{b} \right] + \hat{\beta} \left[\sum_{i=1}^b \sum_{j=1}^{\nu} (x_{ij} - \bar{x})^2 \right]$

$$= \sum_{i=1}^b \sum_{j=1}^{\nu} (y_{ij} - \bar{y}) (x_{ij} - \bar{x})$$

अथवा

$$\hat{\beta} = \frac{\sum_{i=1}^b \sum_{j=1}^{\nu} (y_{ij} - \bar{y}) (x_{ij} - \bar{x}) - \frac{1}{\nu} \sum_{i=1}^b \left(Y_i - \frac{Y}{b} \right) \left(X_i - \frac{X}{b} \right)}{\sum_{i=1}^b \sum_{j=1}^{\nu} (x_{ij} - \bar{x})^2 - \frac{1}{\nu} \sum_{i=1}^b \left(X_i - \frac{X}{b} \right)^2} \quad (257)$$

$$= \frac{\left[\sum_{i=1}^b \sum_{j=1}^{\nu} y_{ij} x_{ij} - \frac{X}{b} \frac{Y}{\nu} \right] - \frac{1}{\nu} \left[\sum_{i=1}^b Y_i X_i - \frac{Y}{b} \frac{X}{\nu} \right]}{\left[\sum_{i=1}^b \sum_{j=1}^{\nu} x_{ij}^2 - \frac{X^2}{b\nu} \right] - \frac{1}{\nu} \left[\sum_{i=1}^b X_i^2 - \frac{X^2}{b} \right]}$$

§ २५४ बिना परिकल्पना के समाश्रयण प्रतिरूप के प्राक्कलों का प्राक्कलन

ये प्राक्कलक तो हमें निराकरणिय परिकल्पना के अंतर्गत प्राप्त हुए। यदि इस परिकल्पना के बिना समीकरण (251) के आधार पर हम μ , t_j , b और β

का प्राक्कलन करें और इनको क्रमशः $\check{\mu}$, $\check{\sigma}$, तथा $\check{\beta}$ से सूचित करें तो इनके लिए हमें निम्नलिखित समीकरण प्राप्त होते हैं।

$$(क) \quad b \nu \check{\mu} = Y$$

अथवा
$$\check{\mu} = \frac{Y}{b\nu} \quad (25.8)$$

$$(ख) \quad r(\check{\mu} + \check{\sigma}) + \check{\beta} \left(X_i - \frac{X}{b} \right) = y_i$$

अथवा
$$\check{\sigma} + \frac{\check{\beta}}{r} \left(X_i - \frac{X}{b} \right) = \frac{1}{r} \left(Y_i - \frac{Y}{b} \right) \quad (25.9)$$

$$(ग) \quad b(\check{\mu} + \check{\sigma}) + \check{\beta} \left(X_j - \frac{X}{\nu} \right) = Y_j$$

अथवा
$$\check{\sigma} + \frac{\check{\beta}}{b} \left(X_j - \frac{X}{\nu} \right) = \frac{1}{b} \left(Y_j - \frac{Y}{\nu} \right) \quad (25.10)$$

$$(घ) \quad \sum_{i=1}^b \check{\sigma} \left(X_i - \frac{X}{b} \right) + \sum_{j=1}^{\nu} \check{\sigma} \left(X_j - \frac{X}{\nu} \right) + \check{\beta} \sum_{i=1}^b \sum_{j=1}^{\nu} (x_{ij} - \bar{x})^2$$

$$= \sum_{i=1}^b \sum_{j=1}^{\nu} (Y_{ij} - \bar{y}) (x_{ij} - \bar{x})$$

अथवा
$$\check{\beta} \left\{ \sum_{i=1}^b \sum_{j=1}^{\nu} (x_{ij} - \bar{x})^2 - \frac{1}{b} \sum_{j=1}^{\nu} \left(X_j - \frac{X}{\nu} \right)^2 - \frac{1}{\nu} \sum_{i=1}^b \left(X_i - \frac{X}{b} \right)^2 \right\}$$

$$= \sum_{i=1}^b \sum_{j=1}^{\nu} (Y_{ij} - \bar{y})^2 (x_{ij} - \bar{x}) - \frac{1}{b} \sum_{j=1}^{\nu} \left(X_j - \frac{X}{\nu} \right) \left(Y_j - \frac{Y}{\nu} \right)$$

$$- \frac{1}{\nu} \sum_{i=1}^b \left(Y_i - \frac{Y}{b} \right) \left(X_i - \frac{X}{b} \right)$$

अथवा
$$\check{\beta} \left[\left\{ \sum_{i=1}^b \sum_{j=1}^{\nu} x_{ij}^2 - \frac{X^2}{b\nu} \right\} - \frac{1}{b} \left\{ \sum_{j=1}^{\nu} X_j^2 - \frac{X^2}{\nu} \right\} - \frac{1}{\nu} \left\{ \sum_{i=1}^b X_i^2 - \frac{X^2}{b} \right\} \right]$$

$$= \left[\left\{ \sum_{i=1}^b \sum_{j=1}^v y_{ij} x_{ij} - \frac{Y \cdot X}{bv} \right\} - \frac{1}{b} \left\{ \sum_{j=1}^v Y_j X_j - \frac{Y \cdot X}{v} \right\} \right. \\ \left. - \frac{1}{v} \left\{ \sum_{i=1}^b X_i Y_i - \frac{Y \cdot X}{b} \right\} \right] \quad (25 \text{ II})$$

इन परिकल्पना के लिए हम एक प्रसरण-सहप्रसरण सारणी को सहायता ले सकते हैं जो पृष्ठ ३५२ पर दी हुई है। जिस प्रकार चर का x प्रसरण $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ होता है उसी प्रकार $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ को x और y का सह प्रसरण कहते हैं।

यदि X और Y यादृच्छिक चर हों तो X और Y का सहप्रसरण $= E (X - m_1)(Y - m_2)$

जहाँ m_1 और m_2 क्रमशः X और Y के प्रत्याशित मान हैं।

यह आसानी से देखा जा सकता है कि

$$\hat{\beta} = \frac{S_{yz} - B_{yz}}{S_{xx} - B_{xx}} = \frac{T_{yz} + E_{yz}}{T_{xx} + E_{xx}}$$

$$\text{और } \check{\beta} = \frac{S_{yx} - B_{yx} - T_{yz}}{S_{xx} - B_{xx} - E_{yz}} = \frac{E_{yx}}{E_{xx}}$$

§ २५.५ उपचार वर्ग-योग

यदि हम प्रतिदर्श प्रेक्षणों में समीकरण (25 I) के प्रतिरूप का भासजन (fitting) करें तो त्रुटि-वर्ग योग निम्नलिखित होगा

$$R_0^2 = \sum_{i=1}^b \sum_{j=1}^v \left[y_{ij} - \hat{\mu} - \hat{b}_i - \hat{t}_j - \hat{\beta} \left(x_{ij} - \frac{X}{bv} \right) \right]^2 \\ = \sum_{i=1}^b \sum_{j=1}^v \left[\left\{ \left(Y_{ij} - \frac{Y}{bv} \right) - \frac{1}{v} \left(Y_i - \frac{Y}{b} \right) \right\} \right]^2$$

सारणी सख्या 25 1
प्रकरण-सहस्ररूप सारणी

विवरण का उपगम	स्वातंत्र्य सख्या	Y^2	XY	X^2
(1)	(2)	(3)	(4)	(5)
आंशिक	$b-1$	$B_{yy} = \frac{1}{v} \sum_{j=1}^b y_j^2 - \frac{Y^2}{bv}$	$B_{xy} = \frac{1}{v} \sum_{j=1}^b y_j x_j - \frac{Y X}{bv}$	$B_{xx} = \frac{1}{v} \sum_{j=1}^b x_j^2 - \frac{X^2}{bv}$
उपचार	$v-1$	$T_{yy} = \frac{1}{b} \sum_{j=1}^v y_j^2 - \frac{Y^2}{bv}$	$T_{xy} = \frac{1}{b} \sum_{j=1}^v y_j x_j - \frac{Y X}{bv}$	$T_{xx} = \frac{1}{b} \sum_{j=1}^v x_j^2 - \frac{X^2}{bv}$
मूटि	$(b-1)(v-1)$	E_{yy}	E_{xy}	E_{xx}
कुल	$bv-1$	$S_{yy} = \sum_{j=1}^b \sum_{i=1}^v y_{ij}^2 - \frac{Y^2}{bv}$	$S_{xy} = \sum_{j=1}^b \sum_{i=1}^v y_{ij} x_{ij} - \frac{Y X}{bv}$	$S_{xx} = \sum_{j=1}^b \sum_{i=1}^v x_{ij}^2 - \frac{X^2}{bv}$

$$-\frac{1}{b} \left(Y_j - \frac{Y \cdot}{v} \right) \left\{ \left(x_j - \frac{X}{bv} \right) - \frac{1}{v} \left(X_i - \frac{X}{b} \right) - \frac{1}{b} \left(X_j - \frac{X}{v} \right) \right\}^2$$

(के लिए समीकरण (क), (ख), (ग) और (घ))

$$\therefore R_0^2 = \sum_{i=1}^b \sum_{j=1}^v \left[\left(y_{ij} - \frac{Y_{i \cdot}}{v} - \frac{Y_{\cdot j}}{b} + \frac{Y \cdot}{bv} \right) - \beta \left(x_{ij} - \frac{X_i}{v} - \frac{X_j}{b} + \frac{X}{bv} \right) \right]^2$$

$$\begin{aligned} & \left(x_{ij} - \frac{X_i}{v} - \frac{X_j}{b} + \frac{X}{bv} \right) + \beta^2 \left(x_{ij} - \frac{X_i}{v} - \frac{X_j}{b} + \frac{X}{bv} \right)^2 \Big] \\ & = E_{yy} - 2 \frac{E_{yz}}{E_{xx}} E_{yx} + \left(\frac{E_{zz}}{E_{xx}} \right)^2 E_{xx} = E_{yy} - \frac{E_{yz}^2}{E_{xx}} E_{yy} - \beta^2 E_{yx} \quad (25.12) \end{aligned}$$

इसी प्रकार समीकरण (25.4) के प्रतिरूप के आसजन करने पर वृत्ति निम्नलिखित होगी

$$R_1^2 = E_{yy} + T_{yy} - \frac{(E_{yz} + T_{yz})^2}{(E_{xx} + T_{xx})} \quad \dots (25.13)$$

$$= E_{yy} + T_{yy} - \hat{\beta} (E_{yz} + T_{yz})$$

इन दोनों वृत्तियों का अंतर हमें उपचार वर्ग-योग देता है।** क्योंकि उपचारो के प्रभाव यदि वास्तव में समान होते तो R_0^2 और R_1^2 के प्रत्याशित मान समान ही होते। इनका अंतर केवल उपचारोके वर्ग-योग के R_1^2 में शामिल हो जाने के कारण है। इस तरह

$$\begin{aligned} \text{उपचार वर्ग-योग} &= R_1^2 - R_0^2 \\ &= \{E_{yy} + T_{yy} - \hat{\beta} (E_{yz} + T_{yz})\} - \{E_{yy} - \beta^2 E_{yx}\} \\ &= T_{yy} - \hat{\beta} (E_{yz} + T_{yz}) + \beta^2 E_{yx} \end{aligned}$$

**उपचार वर्ग-योग प्राप्त करने की यह विधि सामान्य (general) है। पिछले प्रयोगो के विश्लेषण में भी उपचार वर्ग-योग को इस विधि से प्राप्त किया जा सकता था परंतु वहाँ की हुई विधि अधिक सरल होने के कारण इस सामान्य विधि का वर्णन पिछले अध्यायो में नहीं किया गया था।

२५.६ परिकल्पनाओं के परीक्षण

इसलिए यदि हम इस निराकरण योग्य परिकल्पना की परीक्षा करना चाहते हैं कि सब उपचारों के प्रभाव समान हैं तो हमें उपचार-वर्ग माध्य और त्रुटि-वर्ग माध्य के अनुपात का कलन करना चाहिए। यदि यह अनुपात $F_{p-1, b, p-1, b}$ बटन के एक पूर्व निश्चित प्रतिशत बिंदु से अधिक हो तो हम निराकरण योग्य परिकल्पना को अस्वीकार कर देंगे।

यदि परिकल्पना अस्वीकृत होती है तो हमारी चेष्टा यह जानने की होती है कि कौन-कौन से उपचारों के प्रभावों के अंतर अर्थ-पूर्ण हैं। उपचार प्रभाव t_1 और t_2 के अंतर का प्राक्कलन निम्नलिखित है।

$$\hat{t}_1 - \hat{t}_2 = \frac{1}{b} [(Y_{1j} - Y_{2j}) - \hat{\beta}(X_{1j} - X_{2j})] \quad \dots\dots(25.15)$$

इस प्राक्कलन का प्रसरण निम्नलिखित है।

$$\frac{2\sigma^2}{b} + \frac{\sigma^2}{b} \frac{(X_{1j} - X_{2j})^2}{E_{xx}} \quad \dots\dots(25.16)$$

इस प्रकार प्रत्येक उपचार युग्म के अंतर के प्राक्कलन का प्रसरण भिन्न होता है।

आइए, अब जो भी कुछ गणित हमने सहप्रसरण के विश्लेषण के सबंध में सीखा है उसका उपयोग एक उदाहरण में करके उससे अधिक परिचित हो जायें।

§ २५.७ उदाहरण

तीन प्रकार की खादें हैं। इनका प्रभाव गेहूँ की उपज पर क्या है यह जानने के लिए एक यादृच्छिकीकृत ब्लॉक अभिकल्पना का उपयोग किया गया। इस प्रयोग में कुल पाँच ब्लॉक थे। प्रत्येक ब्लॉक में तीन बराबर बराबर क्षेत्रफल के प्लॉट थे। इन तीन प्लॉटों में तीन प्रकार की खादों का प्रयोग किया गया। किस प्लॉट में कौन सी खाद का उपयोग किया जाय यह यादृच्छिकीकरण द्वारा निश्चय किया गया। इन विभिन्न खाद पाने वाले प्लॉटों में उपज की तुलना करके यह पता चल सकता है कि इन खादों के प्रभाव में कोई विशेष अंतर है या नहीं।

परंतु इस प्रयोग में ब्लॉक-प्रभाव, खाद-प्रभाव और प्लॉट-प्रभाव के अतिरिक्त विचरण का एक और उद्गम है और वह है पौधों की संख्या। यद्यपि तीनों प्लॉटों में क्षेत्रफल बराबर है परंतु गेहूँ बोनो का तरीका ऐसा हो सकता है कि इन प्लॉटों में पौधों की संख्या भिन्न-भिन्न हो। यह स्पष्ट है कि इस संख्या के अधिक या कम होने का

प्रभाव कुल उपज को बढ़ाने अथवा घटाने में सहायता पहुँचायगा। फिर भी यह आवश्यक नहीं है कि उपज पीधो की संख्या के अनुपात में ही हो। यद्यपि इस उद्गम से उत्पन्न विचरण को भी श्रुटि का एक भाग मानकर प्रयोग का विश्लेषण किया जा सकता है तथापि इस प्रकार के विश्लेषण में प्राक्कलनों का प्रसरण अधिक होगा तथा निराकरणाय परिकल्पना का परीक्षण सामर्थ्यवान (powerful) नहीं होगा। यदि इस उद्गम से उत्पन्न विचरण को हम सह प्रसरण विश्लेषण द्वारा हटा सकें तो परीक्षण की सामर्थ्य (power) बढ़ जायगी।

इसके लिए ब्लॉक i के जिस प्लॉट में j -वीं खाद का प्रयोग हुआ है उसको (ij) से सूचित करेंगे। (ij) प्लॉट की उपज को हम Y_{ij} तथा उसमें पीधो की संख्या को हम x_{ij} से सूचित करेंगे।

निराकरणाय परिकल्पना H_0 —खादों के प्रभाव समान हैं।

वैकल्पिक परिकल्पना H_1 —खादों के प्रभाव समान नहीं हैं।

§ २५७१ प्रेक्षण

प्रयोग के फल नीचे की सारणी में दिये हुए हैं।

सारणी संख्या 252

उपचार ब्लॉक i \ j	Y_{ij}				x_{ij}			
	1	2	3	कुल Y_i	1	2	3	कुल X_i
1	5	7	11	23	70	100	143	313
2	6	8	9	23	91	108	114	313
3	7	6	6	19	102	82	72	256
4	■	8	9	23	85	111	118	314
5	8	7	10	25	114	94	129	337
कुल Y, X_i	32	36	45	$\frac{113}{=Y}$	462	495	576	$\frac{1,533}{=X}$

§ २५.७ २ विह्लेपण

$$\left\{ \begin{array}{l} \frac{\lambda^2}{5 \times 3} = 156\,672\,60 \\ \frac{X Y}{5 \times 3} = 11\,548\,60 \\ \frac{Y^2}{5 \times 3} = 851\,27 \end{array} \right.$$

$$\left\{ \begin{array}{l} \sum_{i=1}^5 \sum_{j=1}^3 x_{ij}^2 = 162,545\,00 \\ \sum_{i=1}^5 \sum_{j=1}^3 x_{ij} y_{ij} = 12\,017\,00 \\ \sum_{i=1}^5 \sum_{j=1}^3 y_{ij}^2 = 891\,00 \end{array} \right.$$

$$\left\{ \begin{array}{l} \frac{1}{5} \sum_{i=1}^5 X_i^2 = 157,879\,67 \\ \frac{1}{5} \sum_{i=1}^5 X_i Y_i = 11\,636\,33 \\ \frac{1}{5} \sum_{i=1}^5 Y_i^2 = 857\,67 \end{array} \right.$$

$$\frac{1}{5} \sum_{i=1}^5 X_i^2 = 158,049\,00$$

$$\frac{1}{5} \sum_{i=1}^5 X_i Y_i = 11,704\,80$$

$$\frac{1}{5} \sum_{j=1}^3 Y_j^2 = 869\,00$$

सारणी संख्या 253

प्रसरण और सह-प्रसरण विश्लेषण सारणी

विचरण का उद्गम	स्वातंत्र्य संख्या	y^2	xy	x^2
(1)	(2)	(3)	(4)	(5)
श्लोक	4	$B_{yy} = 6.40$	$B_{yz} = 87.73$	$B_{zz} = 1207.07$
उपचार	2	$T_{yy} = 17.73$	$T_{yz} = 156.20$	$T_{zz} = 1376.40$
श्रुति	8	$E_{yy} = 15.60$	$E_{yz} = 224.47$	$E_{zz} = 3289.93$
कुल	14	$S_{yy} = 39.73$	$S_{yz} = 468.40$	$S_{zz} = 5873.40$

यदि सहकारी चर के प्रभाव की उपेक्षा कर दी जाती तो उपचारों की तुलना के लिए हमारा निकष $\frac{T_{yz}/2}{E_{yy}/8} = F$ होता जिसका बटन परिकल्पना के सत्य होने पर $F_{2,9}$ होता। इस प्रयोग में F का मान 4.55 है जो $F_{2,9}$ के पाँच प्रतिशत बिंदु 4.46 से अधिक है। (देखाएँ सारणी संख्या 11.1) इसलिए हम निराकरणाय परिकल्पना को अस्वीकार कर देते। परंतु यह बहुत संभव है कि इस अस्वीकृति का कारण खाद के प्रभावों में अंतर नहीं बल्कि बीघों की संख्या में अंतर हो। यह भी संभव है कि खाद के प्रभावों का अंतर 1 प्रतिशत बिंदु पर भी अर्थपूर्ण हो। आइए अब हम बीघों की संख्या के प्रभाव को सहप्रसरण विश्लेषण द्वारा हटाकर देखें कि हमारे ऊपर के निष्कर्ष में कुछ अंतर पड़ता है या नहीं।

$$\hat{\beta} = \frac{E_{yz}}{E_{zz}} = \frac{224.47}{3289.93} = 0.06823$$

$$\hat{\beta} E_{yy} = 0.06823 \times 224.47 = 15.32$$

$$E'_{yy} = E_{yy} - T_{yy} = 33.33$$

$$E'_{yz} = E_{yz} - T_{yz} = 380.67$$

$$E'_{zz} = E_{zz} - T_{zz} = 4,666.33$$

$$\therefore \hat{\beta} = \frac{E'_{yz}}{E'_{zz}} = 0.08158$$

$$\text{तथा } \hat{\beta} \times E'_{zz} = 0.08158 \times 380.67 \\ = 31.06$$

$$\text{त्रुटि वर्ग योग} = E'_{yy} - \hat{\beta} E'_{yz} = 15.60 - 15.32 \\ = 0.28$$

क्योंकि E'_{yy} की स्वातन्त्र्य सख्या 8 तथा $\hat{\beta} E'_{yz}$ की स्वातन्त्र्य सख्या 1 है इसलिए $E'_{yy} - \hat{\beta} E'_{yz}$ की स्वातन्त्र्य सख्या 7 है।

$$(\text{उपचार} + \text{त्रुटि}) \text{ वर्ग-योग} = E'_{yy} - \hat{\beta} E'_{yz} = 33.33 - 31.06 \\ = 2.27$$

$$\therefore \text{उपचार वर्ग-योग} = 2.27 - 0.28 = 1.99$$

क्योंकि E'_{zz} की स्वातन्त्र्य सख्या 10 है तथा $\hat{\beta} E'_{yz}$ की स्वातन्त्र्य सख्या 1 है इसलिए $E'_{zz} - \hat{\beta} E'_{yz}$ की स्वातन्त्र्य सख्या 9 है।

सारणी संख्या 254

पौधों की संख्या के प्रभाव को हटाने के बाद उपचार-प्रभाव की जाँच

उद्गम	स्वातन्त्र्य सख्या	वर्ग-योग	वर्ग-माध्य	अनुपात F
(1)	(2)	(3)	(4)	
उपचार	2	1.99	1.00	25.00
त्रुटि	7	$E'_{yy} - \hat{\beta} E'_{yz} = 0.28$	0.04	
उपचार + त्रुटि	9	$E'_{yy} - \hat{\beta} E'_{yz} = 2.27$		

निकष F का यह मान एक प्रतिशत स्तर पर भी अर्थपूर्ण है। जब कि सहकारी चर की उपेक्षा करने पर प्रेक्षण फल 1 प्रतिशत स्तर पर अर्थहीन है। इससे यह मालूम होता है कि सहकारी चर का प्रभाव हटा देने से हमारा परीक्षण अधिक शक्तिशाली हो सकता है।

प्रयोग-अभिकल्पनाएँ अन्य भी अनेक प्रकार की होती हैं परंतु उनका विवरण देने का न तो इस पुस्तक में स्थान है और न यह आवश्यक ही है। अतः प्रयोग-अभिकल्पना के विवरण को हम यहीं समाप्त करते हैं।

भाग ६
प्रतिदर्श सर्वेक्षण
Sample Survey

अध्याय २६

प्रतिदर्श-सर्वेक्षण के साधारण सिद्धांत

General Principles of Sample Survey

सरल यादृच्छिक प्रतिचयन

Simple Random Sampling

६ २६.१ योजना के लिए सर्वेक्षण की आवश्यकता

किसी भी योजना को बनाने के पूर्व कुछ आँकड़ों की आवश्यकता होती है। नान कीजिए कि उत्तर प्रदेश सरकार का उद्देश्य १४ वर्ष से छोटे सब बालक-बालिकाओं को निःशुल्क शिक्षा देना है। इसके लिए यह निश्चित करना होगा कि किस-किस स्थान पर कितने स्कूल खोले जायें और उनमें कितने अध्यापक रखे जायें। इसके पूर्व कि सरकार इस प्रकार का कोई निश्चय करे उसे कदाचित् निम्नलिखित बातों का ध्यान रखना होगा।

(१) १४ वर्ष से कम के बालक-बालिकाओं की संख्या कितनी है और वह किस गति से बढ़ रही है। यदि सरकार की इस बारे में कोई भी नीति है कि एक स्कूल में अधिक से अधिक कितने विद्यार्थियों को पढ़ना चाहिए और विद्यार्थियों और शिक्षकों की संख्या में क्या अनुपात रहना चाहिए तो सरकार को साधारण रूप में यह ज्ञान ही ज्ञापना कि इस योजना के लिए कितने स्कूल और कितने शिक्षकों की आवश्यकता है।

(२) वर्तमान स्थिति में उत्तर-प्रदेश में कितने स्कूल हैं—उनमें कितने विद्यार्थी और शिक्षक हैं। यदि सरकार का शिक्षक-विद्यार्थी अनुपात अथवा एक स्कूल में विद्यार्थियों की संख्या के बारे में कोई निश्चित मत नहीं है तो इस मत के स्थिर करने में यह सूचना उपयोगी सिद्ध हो सकती है। इसके अतिरिक्त इससे यह पता चलेगा कि वर्तमान स्कूलों के अतिरिक्त कितने नये स्कूलों की स्थापना करना आवश्यक है।

(३) सरकार को विभिन्न स्थानों पर जन-संख्या का वितरण और एक स्थान से दूसरे स्थान तक जाने के लिए सड़कों इत्यादि का ज्ञान यह निश्चय करने के लिए आवश्यक है कि स्कूल कहाँ खोले जायें।

(४) सरकार को उन पढ़े-लिखे लोगों की सख्या का भी ज्ञान होना चाहिए जो इन स्कूलों में शिक्षक का पद ग्रहण करने योग्य है और शिक्षक बनने के लिए राजी है।

हो सकता है कि इसके अलावा और भी अनेक प्रकार की सूचनाओं की आवश्यकता योजना बनाने वालों को हो। यह केवल एक उदाहरण था परंतु आप स्वयं विभिन्न योजनाओं को ध्यान में रखकर यह पता लगा सकते हैं कि हर एक के लिए आंकड़ों की आवश्यकता होती है। यह आंकड़े प्रायः ऐसी समष्टियों से संचय रखते हैं जिनमें कुल इकाइयों की सख्या परिमित (finite) होती है। इन आंकड़ों को प्राप्त करने के लिए बहुधा सर्वेक्षण करना पड़ता है। यद्यपि समष्टि परिमित होती है परंतु प्रायः इकाइयों की सख्या इतनी अधिक होती है कि सर्वेक्षण को समष्टि के एक प्रतिदर्श तक ही सीमित रखना पड़ता है। इस प्रकार के सर्वेक्षण को प्रतिदर्श सर्वेक्षण (sample survey) की सजा दी जाती है।

§ २६.२ सर्वेक्षण में त्रुटियाँ

इस तरह के सर्वेक्षण में दो तरह की त्रुटियाँ होती हैं।

(१) प्रतिचयन त्रुटि (Sampling error) — समष्टि से चुने हुए विभिन्न प्रतिदर्शों द्वारा हमें विभिन्न प्राक्कलक प्राप्त होते हैं जो केवल इसी कारण समष्टि प्राचल से भिन्न होते हैं कि प्रतिदर्श में समष्टि की हर एक इकाई नहीं होती। इस कारण से प्राक्कलन और प्राचल में जो अंतर होता है उसको प्रतिचयन त्रुटि कहते हैं। विभिन्न प्रतिदर्शों के लिए यह त्रुटि भिन्न भिन्न होगी। किसी यादृच्छिक प्रतिचयन विधि के लिए इन त्रुटियों के वर्ग के माध्य को प्राक्कलक की माध्य-वर्ग-त्रुटि (mean square error) कहते हैं। यह किसी विशेष प्रतिचयन विधि और प्राक्कलन विधि से संबंधित त्रुटि का एक माप है।

(२) अ-प्रतिचयन त्रुटि (Non-Sampling error) — सर्वेक्षण में त्रुटि के धीरे भी उद्गम हैं। मान लीजिए कि हमें उत्तर प्रदेश के मध्यवर्गीय परिवारों की औसत आय का प्राक्कलन करना है। प्राक्कलन से पूर्व यह जानना आवश्यक है कि मध्य वर्गीय परिवार में हमारा क्या तात्पर्य है और आय की परिभाषा क्या है। यह भी जानना जरूरी है कि परिवार में किस प्रकार के व्यक्तियों को सम्मिलित माना जायगा। इन सब परिभाषाओं के होते हुए भी बहुत संभव है कि कुछ मध्य-वर्गीय परिवार सर्वेक्षण से छूट जायें और कुछ ऐसे परिवार जो इस परिभाषा को सतुष्ट नहीं करते सर्वेक्षण में गलती से मध्यवर्गीय परिवारों की तरह सम्मिलित कर लिये

जायें। यह भी समझ दें कि कुछ परिवारा की अपनी आय का ठीक पता न हो इसलिए उनसे प्रश्न करके जो आय का अनुमान लगाया जाता है वह वास्तविक आय से भिन्न हो। कुछ कारणों से आय सबकी प्रश्नों का उत्तर जान बूझ कर भी गलत दिया जा सकता है।

अनाज की उपज के सर्वेक्षण में यह पता चलाना होता है कि कितने क्षेत्रफल में अनाज बोया गया है। इस प्रकार के सर्वेक्षण के लिए अनुसंधाता (Investigator) का खेतों पर जाना आवश्यक है और प्रत्येक खेत के लिए—यदि उसका क्षेत्रफल ज्ञात हो—यह पता चलाना आवश्यक है कि उसके क्षेत्रफल के कितने प्रतिशत भाग में अनाज लगा हुआ है। इसके लिए अनुसंधाता अनुमान का आश्रय लेता है। खेत को देखकर वह अनुमान लगाता है कि हमके विज्ञान भाग में अनाज लगा हुआ है। परन्तु स्पष्ट है कि उन दो स्थितियों को छोड़कर जिनमें या तो खेत में अनाज बिल्कुल ही न हो अथवा संपूर्ण क्षेत्र अनाज से ढँका हो, अन्य स्थितियाँ में इस अंदाजे और वास्तविक अनुपात में कुछ न कुछ अंतर अवश्य होता है। यह भी हो सकता है कि कुछ अनुसंधाता ईमानदार न हों और बिना खेत पर गये अपनी दृष्टि से ही इस अनुपात का अनुमान लिख दें। इस प्रकार के अन्य कई कारण हैं जो प्रतिदर्श विशेष से संबंधित नहीं हैं। इस प्रकार की त्रुटियों को अप्रतिचयन त्रुटियाँ कहते हैं।

किसी भी अच्छे सर्वेक्षण का ध्येय इन दोनों प्रकार की त्रुटियों को सीमित रखना होता है। प्रतिचयन त्रुटियों को विशेष प्रतिचयन विधि और प्राक्कलन विधि द्वारा कम किया जा सकता है। यह स्पष्ट है कि यदि प्रतिदर्श में समष्टि की प्रत्येक इकाई ही तो प्रतिचयन त्रुटि शून्य होगी। अप्रतिचयन त्रुटियों को कम करने के लिए अनुसंधाताओं के शिक्षण और नियंत्रण की आवश्यकता है। वे जितने अधिक अनुभवी होंगे और उनपर जितना अधिक नियंत्रण रहेगा उननी ही अप्रतिचयन त्रुटियाँ कम होगी। यह ध्यान देने की बात है कि प्रतिदर्श-परिमाण बढ़ने से प्रतिचयन त्रुटि ती पड़ती है परन्तु अप्रतिचयन त्रुटि बढ़ती है। यह समझ है कि एक छोटे प्रतिदर्श से प्राप्त प्राक्कलन की कुछ त्रुटि पूरी समष्टि से प्राप्त प्राक्कलन की त्रुटि से कम हो।

§ २६३ अन्य उपादान

त्रुटि के अतिरिक्त सर्वेक्षण में और भी कई उपादानों (factors) का विचार रखना पड़ता है। इनमें घन और समय विशेष उल्लेखनीय हैं। किन्हीं भी सर्वेक्षण के लिए एक निश्चित मात्रा से अधिक घन व्यय करना संभव नहीं होता।

जितना अधिक प्रतिदर्श परिमाण होगा उतना ही अधिक धन व्यय करना पड़ेगा। जो धन सर्वेक्षण पर व्यय करना पड़ता है उसे सर्वेक्षण-व्यय (cost of survey) कहते हैं। और यह प्रतिदर्श-परिमाण पर ही नहीं बल्कि प्रतिचयन विधि और प्राक्कलन विधि पर भी निर्भर करता है।

यदि सर्वेक्षण द्वारा आँकड़े बहुत देर में प्राप्त हो तो उनका महत्त्व घट जाता है। उदाहरण के लिए भारत में १९५९ में उत्पन्न खाद्यान्तों के आँकड़ों की आवश्यकता इसलिए पड़ सकती है कि सरकार आयात-निर्यात के द्वारे में कोई निश्चय कर सके। यदि अन्न आवश्यकता से बहुत कम हुआ हो तो लोगों को भूख से बचाने के लिए विदेशों से अन्न मँगाना पड़ेगा। और यदि अन्न आवश्यकता में अधिक हुआ हो तो मशीनों आदि के त्रय के लिए इसको विदेशों में बेचकर विदेशी कय मुद्रा प्राप्त की जा सकती है। परन्तु यदि यह आँकड़े हमें १९६२ तक प्राप्त हो तो उनका महत्त्व समाप्त हो जाता है। क्योंकि यदि अन्न की कमी हुई हो तो उसका असर उस समय तक पड़ ही चुका होगा और आँकड़ा का उपयोग सरकार के आलोचक केवल यह कह सकने के लिए कर सकेंगे कि सरकार को १९६० में अमुक नीति अपनानी चाहिए थी और उसने दूसरी नीति अपना कर गलती की। प्राक्कलनों को थोड़े समय में प्राप्त करने के लिए भी यह आवश्यक है कि प्रतिदर्श बहुत बड़ा न हो।

सर्वेक्षण के सिद्धांतों का अभिप्राय धन समय और अन्य अनुबन्धों के अनुगत एक ऐसी प्रतिचयन विधि और प्राक्कलन विधि को प्राप्त करना है जिसके लिए प्राक्कलन त्रुटि न्यूनतम हो। हम यहाँ केवल प्रतिचयन त्रुटि पर विचार करेंगे क्योंकि अन्य त्रुटियों को कम करने के लिए प्रतिचयन विधि और प्राक्कलन विधि नहीं बरन् शिक्षण, नियंत्रण और अनुभव की आवश्यकता है।

§ २६.४ सरल यादृच्छिक प्रतिचयन (Simple Random Sampling)

यादृच्छिक प्रतिचयन की कई विधियाँ हैं जिनमें से सबसे सरल का नाम सरल यादृच्छिक प्रतिचयन है। मान लीजिए समष्टि में N इकाइयाँ $U_1, U_2, U_3, \dots, U_N$ हैं। इन N इकाइयों में से n परिमाण के कुल $\binom{N}{n}$ अलग-अलग प्रतिदर्श चुने जा सकते हैं। यदि प्रतिदर्श इस प्रकार चुना जाय कि इन सब प्रतिदर्शों के चुने जाने की प्रायिकता $\frac{1}{\binom{N}{n}}$ हो तो इस विधि को सरल यादृच्छिक प्रतिचयन कहते हैं। इसकी विधि यह है कि पहिले तो N इकाइयों में से एक इकाई इस प्रकार चुनी जाय कि सब

इकाइयों के चुने जाने की प्रायिकता समान अर्थात् $\frac{1}{N}$ हो। फिर बाकी बची हुई $(N-1)$ इकाइयों में से एक इकाई इस प्रकार चुनी जाय कि इन बची हुई इकाइयों में से प्रत्येक की चुने जाने की प्रायिकता समान याने $\frac{1}{N-1}$ हो। इसी तरह क्रमशः एक एक करके N इकाइया को इस प्रकार चुना जाय कि प्रत्येक चुनाव में बाकी बची हुई इकाइयों में से प्रत्येक इकाई के चुने जाने की प्रायिकता बराबर रहे।

§ २६५ प्राक्कलन

मान लीजिए हम किसी विशेष चर x के औसत मान \bar{X} का प्राक्कलन करना चाहते हैं। यदि i -वीं इकाई U_i के लिए इस चर का मान X_i है तो

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \quad \text{--- (26.1)}$$

हम i -वीं चुनी हुई इकाई के लिए x के मान को x_i से सूचित करेंगे। x_1, x_2, \dots, x_n सभी यादृच्छिक चर हैं जो प्रत्येक मान $X_j, j=1, 2, \dots, N$ को समान प्रायिकता $\frac{1}{N}$ से ग्रहण करते हैं। यदि हम प्रतिदर्श माध्य को \bar{x} से सूचित करें तो

$$E(\bar{x}) = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right]$$

$$= \frac{1}{n} \sum_{i=1}^n E(x_i)$$

$$\text{परन्तु } E(x_i) = \sum_{j=1}^N X_j \frac{1}{N}$$

$$= \bar{X}$$

$$E(\bar{x}) = \frac{1}{n} \sum_{i=1}^n \bar{X} = \bar{X}$$

इस प्रकार हम देखते हैं कि \bar{X} का एक अनमिथता प्राक्कलक \bar{x} है। किसी दूसरी प्रतिचयन विधि से तुलना करने के पूर्व यह जानना आवश्यक है कि इस प्राक्कलक का प्रसरण कितना है।

§ २६.६ प्राक्कलक का प्रसरण

$$\begin{aligned} V(\bar{x}) &= E(\bar{x} - \bar{X})^2 \\ &= E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})\right]^2 \\ &= \frac{1}{n^2} \sum_{i=1}^n E(x_i - \bar{X})^2 + \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i} E(x_i - \bar{X})(x_j - \bar{X}) \end{aligned}$$

यह स्पष्ट है कि ऊपर दी हुई प्रतिचयन विधि के लिए

$$E(x_i - \bar{X})^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{X})^2$$

$$\begin{aligned} \text{तथा } E(x_i - \bar{X})(x_j - \bar{X}) &= \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j \neq i} (x_i - \bar{X})(x_j - \bar{X}) \\ &= \frac{1}{N(N-1)} \sum_{i=1}^N (x_i - \bar{X}) \sum_{j \neq i} (x_j - \bar{X}) \\ &= \frac{-1}{N(N-1)} \sum_{i=1}^N (x_i - \bar{X})^2 \end{aligned}$$

$$\text{क्योंकि } \sum_{j \neq i} (x_j - \bar{X}) = \sum_{j=1}^N (x_j - \bar{X}) - (x_i - \bar{X})$$

$$\text{और } \sum_{j=1}^N (x_j - \bar{X}) = 0$$

$$\begin{aligned} \therefore V(\bar{x}) &= \frac{1}{n^2} \left[\frac{n}{N} \sum_{i=1}^N (x_i - \bar{X})^2 - \frac{n(n-1)}{N(N-1)} \sum_{i=1}^N (x_i - \bar{X})^2 \right] \\ &= \frac{N-n}{Nn} \frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N-1} \\ &= \frac{N-n}{Nn} S^2 \quad \dots\dots(26.2) \end{aligned}$$

$$\text{जहाँ } S^2 = \frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N-1} \quad \dots\dots(26.3)$$

यदि प्रतिदर्श परिमाण n यथेष्ट बड़ा हो तो \bar{x} का वृत्त प्रायः प्रसामान्य होगा। यदि हम इसके मानक विचलन का प्राक्कलन कर सकें तो समष्टि प्राचल \bar{X} के लिए विस्वास्य-अंतराल का प्राक्कलन भी किया जा सकता है। हम नीचे $V(\bar{x})$ का प्राक्कलन मालूम करेंगे और उसके वर्गमूल का उपयोग मानक विचलन के प्राक्कलन के लिए करेंगे।

§ २६७ प्राक्कलक के प्रसरण का प्राक्कलन

S के समान एक कलन S^2 हम प्रतिदर्श के लिए परिभाषित करते हैं

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \dots (264)$$

यह सिद्ध करना आवश्यक सरल है कि S^2 का एक अनभिन्नत प्राक्कलक S^2 है।

$$\begin{aligned} E(S^2) &= \frac{1}{n-1} \sum_{i=1}^n E(x_i - \bar{x})^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n E[x_i - X - (\bar{x} - \bar{X})]^2 \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n E(x_i - \bar{X})^2 - n E(\bar{x} - \bar{X})^2 \right] \\ &= \frac{1}{n-1} \left[\frac{n}{N} \sum_{i=1}^N (X_i - \bar{X})^2 - n \frac{N-n}{Nn} \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1} \right] \\ &= \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N(N-1)(n-1)} [n(N-1) - (N-n)] \\ &= \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1} = S^2 \quad \dots (265) \end{aligned}$$

∴ $V(\bar{x})$ का अनभिन्नत प्राक्कलक $\hat{V}(\bar{x}) = \frac{N-n}{Nn} S^2$

हम साधारणतया किसी प्राचल θ के प्राक्कलक को $\hat{\theta}$ से सूचित करेंगे।
यदि हम समष्टि योग $X = \sum_{i=1}^N X_i$ का प्राक्कलन करना चाहें तो स्पष्टतया

$$\hat{X} = N \bar{x} \quad \dots (26.6)$$

$$V(\hat{X}) = N^2 V(\bar{x}) = \frac{N(N-n)}{n} S^2 \quad \dots (26.7)$$

$$\hat{D}(\hat{X}) = \frac{N(N-n)}{n} s^2 \quad \dots (26.8)$$

$\therefore S^2$ का अनभिन्नत प्राक्कलक s^2 है।

§ २६८ अनुपात का प्राक्कलन

ऊपर दिये हुए सूत्रों का उपयोग समष्टि में विशेष गुण वाली इकाइयों के अनुपात के प्राक्कलन के लिए भी किया जा सकता है। उदाहरण के लिए मान लीजिए कि एक नगर में N व्यक्ति हैं जिनमें से N_1 की उम्र १४ वर्ष अथवा उससे कम है। N_2 हमें श्रात नहीं है। हम नगर में १४ वर्ष से कम उम्र वाले व्यक्तियों का अनुपात $P = \frac{N_1}{N}$ जानना चाहते हैं।

मान लीजिए X_i एक चर है जो i वें व्यक्ति की उम्र १४ वर्ष से कम होने पर मान १ ग्रहण करता है अन्यथा मान ०। इस प्रकार नगर के प्रत्येक व्यक्ति के लिए एक चर है। यह आप देख सकते हैं कि $\sum_{i=1}^N X_i = N_1$ और एक n परिमाण के प्रतिदर्श में $n_1 = \sum_{i=1}^n x_i =$ प्रतिदर्श में १४ वर्ष से कम उम्र के व्यक्तियों की संख्या।

$$\therefore \hat{P} = \left(\frac{\hat{N}_1}{N} \right) = \frac{\hat{X}}{N} = \bar{x} = \frac{n_1}{n} = p \quad \dots (26.9)$$

= प्रतिदर्श में १४ वर्ष से कम उम्र के व्यक्तियों का अनुपात

$$\text{इसी प्रकार } V(p) = \frac{N-n}{Nn} \frac{\sum_{i=1}^N x_i^2 - N\bar{X}^2}{N-1}$$

[देखिए समीकरण (26.2) और समीकरण (26.3)]

$$\begin{aligned}
 &= \frac{N-n}{Nn} \frac{N_1 - N \left(\frac{N_1}{N} \right)^2}{N-1} \\
 &= \frac{N-n}{Nn} \frac{NP - NP^2}{N-1} \\
 &= \frac{N-n}{n(N-1)} P(1-P) \qquad (26\ 10)
 \end{aligned}$$

$$\hat{V}(p) = \frac{N-n}{Nn} \frac{np - np^2}{n-1} \qquad (\text{देखिए समीकरण } 26\ 8)$$

$$= \frac{N-n}{N(n-1)} p(1-p) \qquad (26\ 11)$$

उदाहरण —

यदि $N=1,000$

$n=200$

$n_1=80$

$$\text{तो } \hat{P} = p = \frac{80}{200} = \frac{2}{5}$$

$$\begin{aligned}
 \hat{V}(p) &= \frac{1,000-200}{1,000 \times 199} \times \frac{2}{5} \times \frac{3}{5} \\
 &= \frac{24}{24,875}
 \end{aligned}$$

§ २६९ विचरण-गुणांक और प्रतिदर्श परिमाण

यदि किसी प्राक्कलक t का मानक विचलन σ_t और माध्य μ_t हो तो $\frac{\sigma_t}{\mu_t}$ को t का विचरण गुणांक (coefficient of variation) कहते हैं और इसे $CV(t)$ से सूचित करते हैं। बहुधा सर्वेक्षण का उद्देश्य एक निश्चित मूल्य से कम विचरण गुणांक वाला प्राक्कलक प्राप्त करना होता है। सरल यादृच्छिक प्रतिचयन में विचरण गुणांक केवल प्रतिदर्श परिमाण पर ही निर्भर करता है। \bar{x} का विचरण गुणांक $\sqrt{\frac{N-n}{Nn}} \frac{S}{\bar{X}}$ है। यदि हमें समष्टि के लिए $\frac{S}{\bar{X}}$ का अच्छा अनुमान हो

जिसे हम C से सूचित करें और यदि हम यह चाहते हो कि \bar{x} का विचरण गुणांक लगभग α हो तो हम प्रतिदर्श परिमाण n को निम्नलिखित सूत्र द्वारा निश्चित कर सकते हैं—

$$\sqrt{\frac{N-n}{Nn}} C = \alpha$$

$$\text{अथवा } \frac{1}{n} - \frac{1}{N} = \frac{\alpha^2}{C^2}$$

$$\text{अथवा } \frac{1}{n} = \frac{N\alpha^2 + C^2}{NC^2}$$

$$\text{अथवा } n = \frac{NC^2}{N\alpha^2 + C^2}$$

उदाहरण—यदि हमें यह ज्ञात है कि १४ वर्ष से कम उम्र के व्यक्तियों का अनुपात प्रायः ३० प्रतिशत है तो $\bar{X} = 0.3$,

$$S^2 = \frac{NP(1-P)}{N-1} = \frac{N}{N-1} (0.3 \times 0.7) \quad [\text{दिए गए समीकरण 26 10}]$$

यदि N दृश्य रूप से बड़ा हो तो $\frac{N}{N-1}$ को जगह सरलता के लिए 1 रख लेने से कोई विशेष त्रुटि नहीं होगी। इस प्रकार

$$C^2 = \frac{0.3 \times 0.7}{0.3 \times 0.3} = \frac{7}{3}$$

यदि हम p के विचरण गुणांक को 2 प्रतिशत के लगभग चाहते हैं तो

$$\alpha^2 = (0.02)^2 = 0.0004$$

$$\therefore \text{इच्छित प्रतिदर्श परिमाण } n = \frac{7N/3}{0.0004N + 7/3}$$

यदि N बहुत बड़ा हो तो

$$\begin{aligned} n &= \frac{7}{3 \times 0.0004} \\ &= \frac{70000}{12} \\ &= 5834 \end{aligned}$$

स्तरित प्रतिचयन (Stratified Sampling)

§ २७.१ परिचय

सरल यादृच्छिक प्रतिचयन का प्रयोग केवल उस दशा में किया जाता है जब समष्टि के बारे में कोई ज्ञान न हो अथवा यदि कुछ ज्ञान हो भी तो बहुत मामूली सा। समष्टि के बारे में जिस प्रकार का और जितना ज्ञान होता है उसके अनुसार प्रतिचयन विधि में ससोपन करके उसे अधिक श्रद्ध बनाया जा सकता है। इनमें से एक ससोधित विधि समष्टि को कुछ स्तरों में विभाजित करके प्रत्येक में से अलग-अलग सरल यादृच्छिक प्रतिचयन करने की है। इस विधि को स्तरित सरल यादृच्छिक प्रतिचयन (stratified simple random sampling) कहते हैं।

§ २७.२ प्रावकलन

मान लीजिए समष्टि को k स्तरों में विभाजित कर दिया गया है जिसमें से i -वें स्तर को S_i से सूचित किया जाएगा। मान लीजिए कि S_i में कुल N_i इकाइयाँ हैं और इसकी j वी इकाई के लिए x का मान X_{ij} है। इसके अतिरिक्त

$$\sum_{j=1}^{N_i} X_{ij} = X_i$$

$$\sum_{i=1}^k X_i = X \qquad \sum_{i=1}^k N_i = N$$

$$\frac{X}{N} = \bar{X}$$

यदि S_i में से j -वी चुनी हुई इकाई के x के मान को x_{ij} से सूचित किया जाए और यदि i -वें स्तर में से n_i इकाइयाँ चुनी जायँ तो \bar{X}_i का एक अनभिन्नत प्रावकलन

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \text{ है।}$$

$$\begin{aligned}
 \text{इसलिए } E \sum_{i=1}^k N_i \bar{X}_i &= \sum_{i=1}^k N_i \bar{X}_i \\
 &= \sum_{i=1}^k X_i \\
 &= X \quad \dots (27.1)
 \end{aligned}$$

इस प्रकार X का एक अनभिनत प्राक्कलक $\bar{X}_c = \sum_{i=1}^k N_i \bar{X}_i$ है। यह

स्पष्ट है कि \bar{X} का अनभिनत प्राक्कलक $\frac{1}{N} \sum_{i=1}^k N_i \bar{X}_i$ है।

$$\begin{aligned}
 \S 27.3 \text{ प्राक्कलन का प्रसरण } V \left[\sum_{i=1}^k N_i \bar{X}_i \right] &= \sum_{i=1}^k V(N_i \bar{X}_i) \\
 &= \sum_{i=1}^k N_i \left(\frac{N_i - n_i}{n_i} \right) S_i^2 \dots (27.2)
 \end{aligned}$$

$$\text{जहाँ } S_i^2 = \frac{\sum_{j=1}^{N_i} (X_{ij} - \bar{X}_i)^2}{N_i} \dots (27.3)$$

§ 27.4 प्रसरण का प्राक्कलन

$$\hat{V} \left(\sum_{i=1}^k N_i \bar{x}_i \right) = \sum_{i=1}^k \frac{N_i (N_i - n_i)}{n_i} s_i^2 \dots (27.4)$$

$$\text{जहाँ } s_i^2 = \sum_{j=1}^{n_i} \frac{(x_{ij} - \bar{x}_i)^2}{n_i - 1} \dots (27.5)$$

$$\begin{aligned}
 \text{तथा } \hat{V}(\bar{X}) &= \sum_{i=1}^k \frac{N_i (N_i - n_i)}{N^2 n_i} s_i^2 \\
 &= \sum_{i=1}^k \left(\frac{N_i}{N} \right)^2 \left(\frac{1}{n_i} - \frac{1}{N_i} \right) s_i^2 \dots (27.6)
 \end{aligned}$$

§ २७.५ विभिन्न स्तरों में प्रतिदर्श परिमाण का वितरण

२७.५.१ समानुपाती वितरण

अब हमारे सामने समस्या यह है कि कुल प्रतिदर्श परिमाण $n = \sum_{i=1}^k n_i$ के दिये होने पर विभिन्न स्तरों के प्रतिदर्श परिमाण n_i को किस प्रकार निश्चित किया जाय। एक तरीका तो यह है कि प्रतिदर्श परिमाण स्तरों को इवाइया की सख्या के अनुपात में हो। इस प्रकार के वितरण को समानुपाती वितरण (*proportional allocation*) कहते हैं।

समानुपाती वितरण के लिए प्राक्कलक को \hat{X}_{prop} से सूचित किया जायगा।

$$\hat{X}_{prop} = \sum_{i=1}^k N_i \bar{x}_i = \sum_{i=1}^k \frac{N_i}{n} \sum_{j=1}^{n_i} x_{ij} = \frac{N}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij} \quad (27.7)$$

क्योंकि $\frac{N_i}{n_i} = \frac{N}{n} \quad i=1,2, \dots, k$

$$\hat{X}_{prop} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij} = \bar{x}$$

इस प्रकार के वितरण के लिए प्राक्कलक बहुत सरल हो जाता है। इसके लिए

$$\begin{aligned} V(\hat{X}_{prop}) &= \sum_{i=1}^k \frac{N_i(N_i - n_i)}{N^2 n_i} S_i^2 \\ &= \frac{1}{Nn} \sum_{i=1}^k (N_i - n_i) S_i^2 \\ &= \frac{1}{n} \sum_{i=1}^k \frac{N_i}{N} \left(1 - \frac{n_i}{N_i}\right) S_i^2 \\ &= \frac{N-n}{Nn} \sum_{i=1}^k \left(\frac{N_i}{N}\right) S_i^2 \end{aligned} \quad (27.8)$$

$$V(\hat{X}_{prop}^*) = \frac{N-n}{Nn} \sum_{i=1}^k \left(\frac{N_i}{N}\right) S_i^2 \quad (27.9)$$

६ २७५२ अनुकूलतम वितरण

यदि सर्वेक्षण का व्यय प्रत्येक स्तर में केवल प्रतिदश इकाइयों पर निर्भर करता हो और i -वें स्तर में एक इकाई के सर्वेक्षण पर व्यय C_i हो तो संपूर्ण सर्वेक्षण का व्यय फलन C निम्नलिखित होगा।

$$C = \sum_{i=1}^k C_i n_i \quad (27\ 10)$$

हम इस प्रकार के वितरण (n_1, n_2, \dots, n_k) को निर्धारित करना चाहते हैं जिसके लिए प्रसरण दिये होने पर व्यय लघुतम अथवा व्यय C_0 दिये होने पर प्रसरण निम्नतम हो। इस वितरण को मालूम करने के लिए निम्नलिखित विधि का उपयोग करना होगा। सबसे प्रथम हम एक परिमाण Q की परिभाषा देते हैं।

$$Q = V(\hat{X}_{gr}) - \lambda \left[C_0 - \sum_{i=1}^k C_i n_i \right] \quad (27\ 11)$$

$$= \frac{1}{N^2} \sum_{i=1}^k N_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i} \right) S_i^2 - \lambda \left[C_0 - \sum_{i=1}^k C_i n_i \right]$$

Q के निम्नतम मान के लिए n_1, n_2, \dots, n_k का पता निम्नलिखित सूत्रों को हल करने से मिलता है।

$$\frac{\partial Q}{\partial n_i} = 0 \quad i=1, 2, \dots, k \quad (27\ 12)$$

$$\text{अथवा} \quad - \frac{N_i^2 S_i^2}{N^2 n_i^2} + \lambda C_i = 0 \quad i=1, 2, \dots, k$$

$$\therefore n_i = \frac{1}{\sqrt{\lambda}} \frac{W_i S_i}{\sqrt{C_i}} \quad \text{जहाँ } W_i = \frac{N_i}{N} \quad (27\ 13)$$

$$\therefore C_0 = \sum_{i=1}^k n_i C_i = \frac{1}{\sqrt{\lambda}} \sum_{i=1}^k W_i S_i \sqrt{C_i}$$

$$\text{अथवा} \quad \frac{1}{\sqrt{\lambda}} = \frac{C_0}{\sum_{i=1}^k W_i S_i \sqrt{C_i}}$$

$$\therefore n_i = [C_o W_i S_i / \sqrt{C_i}] - \sum_{i=1}^k W_i S_i \sqrt{C_i} \quad i=1,2,\dots,k \dots (27.14)$$

यदि $C_1 = C_2 = \dots = C_k = d$ तो $C_o = nd$

$$\therefore n_i = n \frac{W_i S_i}{\sum_{i=1}^k W_i S_i} \quad (27.15)$$

§ २७६ स्तरण-विधि (method of stratification)

एक समस्या यह है कि यदि समष्टि को k स्तरों में विभाजित करने की स्वतंत्रता हो तो यह विभाजन किस प्रकार किया जाय। यह हम इस प्रकार करना चाहेंगे कि प्राक्कलक का प्रसरण जहाँ तक हो सके कम हो जाय। हम जानते हैं कि

$$V_{ran}(\hat{X}) = \left(\frac{1}{N} - \frac{1}{N}\right) S^2 = \left(\frac{1}{n} - \frac{1}{N}\right) \frac{\sum (X_{ij} - \bar{X})^2}{N-1}$$

{ जहाँ V_{ran} सरल यादृच्छिक प्रतिषयन के लिए प्राक्कलक का प्रसरण है। }

$$= \frac{1}{N-1} \left(\frac{1}{n} - \frac{1}{N}\right) \left[\sum_{i=1}^k (N_i - 1) S_i^2 + \sum_{i=1}^k N_i (\bar{X}_i - \bar{X})^2 \right]$$

यदि N_i और N बहुत बड़े हो तो

$$V_{ran}(\hat{X}) = \frac{1}{n} \sum_{i=1}^k W_i S_i^2 + \sum_{i=1}^k W_i (\bar{X}_i - \bar{X})^2 \dots (27.16)$$

$$\text{जहाँ } W_i = \frac{N_i}{N}$$

$$\text{और } V(\hat{X}_{prop}) = \frac{1}{n} \sum_{i=1}^k W_i S_i^2 \dots (27.17)$$

$$\therefore V_{ran}(\hat{X}) - V(\hat{X}_{prop}) = \frac{1}{n} \sum_{i=1}^k W_i (\bar{X}_i - \bar{X})^2 \dots (27.18)$$

यदि हम समानुपाती वितरण प्राप्त करने का विचार रखते हैं तो हम समष्टि को इस प्रकार स्तरित करना चाहेंगे कि ऊपर लिखित अंतर अधिकतम हो। इसके लिए विभिन्न स्तरों की समष्टियों के माध्यों में अधिक से अधिक अंतर होना चाहिए।

§ २७ ■ सन्निकटन (approximation)

इस प्रकार के अनुकूलतम वितरण और अनुकूलतम स्तरण को तभी प्राप्त किया जा सकता है जब हमें समष्टि के बारे में यथेष्ट जानकारी हो। उदाहरण के लिए अनुकूलतम वितरण में S_i के ज्ञान की आवश्यकता है। परन्तु यह ऐसा समष्टि प्राचल है जिसका ज्ञान सर्वेक्षण के पूर्व नहीं हो सकता। इसके अज्ञात होने की अवस्था में हम समानुपाती वितरण का प्रयोग करके ही सन्तुष्ट हो सकते हैं। यदि हमें S_i के किसी अच्छे प्राक्कलन S'_i का ज्ञान हो तो वितरण इसके आधार पर करने से आशा की जा सकती है कि वितरण अनुकूलतम वितरण से बहुत भिन्न नहीं होगा।

यह भी हो सकता है कि हमें x से अनिष्ट रूप से संबंधित किसी और चर y के लिए S'_i का ज्ञान हो जहाँ

$$S'_i{}^2 = \frac{\sum_{i=1}^N (Y_{ij} - \bar{Y}_i)^2}{N_i - 1}$$

और यह विश्वास हो कि $\frac{S'_i}{S_i}$ लगभग अचर है तो n_i का कलन S'_i के आधार पर किया जा सकता है। इस प्रकार के तरीके को अनुकूलतम परिस्थिति के लिए सन्निकटन कहते हैं। यदि इस सन्निकटन और समानुपाती वितरण में अधिक अंतर न हो तो समानुपाती वितरण का ही उपयोग अधिक अच्छा है क्योंकि इससे प्रसरण में विशेष अंतर नहीं पड़गा जब कि प्राक्कलन बहुत सरल हो जायगा।

इसी प्रकार अनुकूलतम स्तरण के लिए $\sum_{i=1}^k W_i (\bar{Y}_i - \bar{Y})^2$ के मान को महत्तम बनाने की चेष्टा की जा सकती है जहाँ \bar{Y}_i और \bar{Y} के मान ज्ञात हैं। इस प्रकार का स्तरण लगभग अनुकूलतम होगा।

द्वि-चरणी प्रतिचयन (Two-stage sampling)

§ २८.१ प्रतिचयन विधि और व्यय

ऊपर लिखी प्रतिचयन विधियों के लिए यह आवश्यक है कि प्रतिचयन कर्ता के पास सभी इकाइयों की एक सूची हो। बहुधा यह सम्भव नहीं होता। उदाहरण के लिए यदि हम भारतीय किसान परिवारों का प्रतिदर्श चयन करना चाहते हैं तो सब परिवारों की सूची प्राप्त करना लगभग असम्भव होगा। यदि यह सूची हम बनाना चाहें तो सर्वेक्षण से भी अधिक धन और समय इस सूची के बनाने में लग जायगा। इसलिए हमें किमी और प्रकार की प्रतिचयन विधि का आशय लेना पड़ता है। यदि हमारे पास सब किसान परिवारों की सूची हो भी तो सरल यादृच्छिक प्रतिचयन के अवलम्बन से यह बहुत सम्भव है कि प्रत्येक परिवार एक जगह ही गाँव से चुना जाय। भारत में गाँव की कुल संख्या साठे छ लाख से भी अधिक होने के कारण इस बात की सम्भावना बहुत कम है कि हजार दो हजार परिवारों में से कोई दो परिवार भी एक ही गाँव से चुने जायेंगे। इस प्रकार के सर्वेक्षण में एक गाँव से दूसरे गाँव की यात्रा का व्यय कुल सर्वेक्षण व्यय का एक मुख्य भाग बन जायगा। यह बहुत सम्भव है कि इस यात्रा व्यय कम करके इस धन को अधिक परिवारों के सर्वेक्षण में लयाया जाता तो कुल प्रसरण में कमी हो जाती। इस प्रकार के दो कारण जो विशेष कर व्यय के कम करने से सबध रखते हैं हमें उस प्रतिचयन विधि का अवलम्बन करने का सकेत करने हैं जो द्वि-चरणी प्रतिचयन कहलाता है।

§ २८.२ द्वि-चरणी प्रतिचयन विधि

इसमें प्रतिचयन उत्तरोत्तर दो चरणों में किया जाता है। यदि अंतिम इकाइयों की सूची हमारे पास नहीं है अथवा उनके सरल प्रतिचयन में अपव्यय होता है तो हम पहिले इस प्रकार की इकाइयों के कई समूह बना लेते हैं—साधारणतया यह समूह पहिले से ही बने होते हैं और इनके निर्माण की आवश्यकता नहीं पड़ती। प्रतिचयन के पहिले चरण में हम इन समूहों में से कुछ का चयन करते हैं। इस प्रकार ये समूह प्रतिचयन की प्रथम-चरणी इकाइयाँ कहलाते हैं। इसके बाद इन चुनी हुई प्रथम-

चरणी इकाइयों में से प्रत्येक में से कुछ निश्चित सख्या में अंतिम इकाइयों को चुना जाता है। इस कारण ये द्वितीय-चरणी इकाइयाँ कहलाती हैं। उदाहरणार्थ किसान परिवारों के चयन के लिए पहिले भारत में कुछ गाँवों का चयन किया जा सकता है। इन चुने हुए गाँवों में किसान परिवार की सूची तैयार की जा सकती है। इनमें से कुछ परिवार प्रत्येक चुने हुए गाँव में से चयन किये जा सकते हैं।

§ २८.३ सकेत

मान लीजिए समष्टि में N प्रथम-चरणी इकाइयाँ $U_1, U_2, U_3, \dots, U_N$ हैं। i -वीं इकाई U_i में M_i द्वितीय-चरणी इकाइयाँ $U_{i1}, U_{i2}, \dots, U_{iM_i}$ हैं। मान लीजिए U_j के लिए गुण x का मान X_{ij} है।

$$\begin{aligned} \sum_{j=1}^N \sum_{i=1}^{M_j} X_{ij} &= X \\ \sum_{i=1}^N \sum_{j=1}^{M_i} X_{ij} &= X \\ \frac{X}{\sum_{i=1}^N M_i} &= \bar{x} \end{aligned}$$

§ २८.४ प्रतिचयन -

पहिले प्रथम-चरणी इकाइयों में से n परिमाण का एक सरल यादृच्छिक प्रतिदर्श चुनते हैं। चुनी हुई इकाइयों में से i -वीं के गुण x के मान को हम x_i से सूचित करेंगे। इस i -वीं इकाई की कुल M_i इकाइयों में से हम m_i द्वितीय-चरणी इकाइयाँ सरल यादृच्छिक प्रतिचयन द्वारा चुनते हैं। इसकी j वीं चुनी हुई द्वितीय-चरणी इकाई के x गुण के मान को हम x_{ij} से सूचित करेंगे।

§ २८.५ प्राक्कलन

इस द्वितीय-चरणी चयन के लिए $\frac{M_i}{m_i} \sum_{j=1}^{m_i} x_{ij}$ को X_i का प्राक्कलक माना जा सकता है।

$$E_x \left[\frac{M_i}{m_i} \sum_{j=1}^{m_i} x_{ij} \right] = x_i$$

यहाँ हम E_2 द्वारा प्रथम-चरणी इकाई दिये होने पर द्वितीय चरणी इकाइयो पर आश्रित प्राक्कलक के प्रत्याक्षित मान को सूचित करते हैं।

$$E_1 \frac{N}{n} \sum_{i=1}^n x_i = X$$

$$\therefore \hat{X} = \frac{N}{n} \sum_{j=1}^m \frac{M_j}{m_j} \sum_{i=1}^{m_j} x_{ij}$$

$$= \frac{N}{n} \sum_{j=1}^m M_j \bar{x}_j \quad (28.1)$$

§ २८६ प्राक्कलक प्रसरण

$$V(\hat{X}) = E_1 E_2 (\hat{X})^2 - X^2$$

$$= E_2 [V_2(\hat{X}) + \{E_2(\hat{X})\}^2] - X^2$$

$$= E_1 V_2(\hat{X}) + [E_1 \{E_2(\hat{X})^2\} - X^2]$$

$$= E_1 V_2(\hat{X}) + V_1 E_2(\hat{X})$$

$$E_2(\hat{X}) = \frac{N}{n} \sum_{i=1}^n x_i$$

$$\therefore V_1 E_2(\hat{X}) = \frac{N(N-n)}{n} \frac{\sum_{i=1}^n \left(x_i - \frac{X}{N}\right)^2}{N-1} \quad (28.2)$$

$$\text{और } V_2(\hat{X}) = \frac{N^2}{n^2} \sum_{j=1}^m \frac{M_j(M_j - m_j)}{m_j} \frac{\sum_{i=1}^{m_j} \left(x_{ij} - \frac{X_j}{M_j}\right)^2}{M_j - 1}$$

$$\therefore E_1 V_2(\hat{X}) = \frac{N}{n} \sum_{j=1}^m \frac{M_j(M - m_j)}{m_j} \frac{\sum_{i=1}^{m_j} \left(x_{ij} - \frac{X_j}{M_j}\right)^2}{M_j - 1} \quad (28.3)$$

हम $\frac{N}{N-1} \sum_{i=1}^N \left(X_i - \frac{X}{N} \right)^2$ को $M^2 S_b^2$ और $\frac{M_i}{M_i-1} \sum_{j=1}^{M_i} \left(X_{ij} - \frac{X_i}{M_i} \right)^2$ को

$$S_i^2, \text{ द्वारा सूचित करेंगे जहाँ } M = \frac{1}{N} \sum_{i=1}^N M_i$$

$$\therefore V(\hat{X}) = \frac{N(N-n)}{n} M^2 S_b^2 + \frac{N}{n} \sum_{i=1}^N \frac{M_i(M_i-m_i)}{m_i} S_i^2 \quad (284)$$

§ २८७ प्रसरण का प्राक्कलन

यदि हम द्वितीय चरणी इकाइया के आधार पर s_i^2 से S_i^2 का प्राक्कलन करें तो

$$S_i^2 = \frac{1}{m_i-1} \sum_{j=1}^{m_i} \left(x_{ij} - \frac{1}{m_i} \sum_{j=1}^{m_i} x_{ij} \right)^2 \quad (285)$$

तथा (284) के दूसरे भाग का प्राक्कलक स्पष्टतया $\frac{N^2}{n^2} \sum_{i=1}^n \frac{M_i(M_i-m_i)}{m_i}$

s_i^2 है। इसी प्रकार प्रथम भाग का प्राक्कलन भी प्राप्त किया जा सकता है।

$$\begin{aligned} E \sum_{i=1}^n (N\hat{X}_i - \hat{X})^2 &= N^2 n V(\hat{X}_i) - n V(\hat{X}) \\ &= \frac{N^2(n-1)}{N-1} \sum_{i=1}^n (X_i - \bar{X})^2 + N(n-1) \sum_{i=1}^n \frac{M_i(M_i-m_i)}{m_i} S_i^2 \end{aligned}$$

इसलिए प्रथम भाग का प्राक्कलन निम्नलिखित है

$$\frac{N(N-n)}{n} \left[\frac{\sum_{i=1}^n \left(\hat{X}_i - \frac{\hat{X}}{N} \right)^2}{n-1} - \frac{1}{n} \sum_{i=1}^n \frac{M_i(M_i-m_i)}{m_i} S_i^2 \right] \quad (285)$$

$$\begin{aligned} \text{इस प्रकार } \hat{V}(\hat{X}) &= \frac{N(N-n)}{n} \sum_{i=1}^n \left(\hat{X}_i - \frac{\hat{X}}{N} \right)^2 \\ &+ \frac{N}{n} \sum_{i=1}^n \frac{M_i(M_i-m_i)}{m_i} S_i^2, \quad \dots (286) \end{aligned}$$

यदि प्रत्येक प्रथम-चरणी इकाई में M इकाइयाँ हा जिनमे से m चुनी जायें तो

$$\hat{X} = \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m x_{ij} \quad (28.7)$$

$$V(\hat{X}) = \frac{N-n}{Nn} S_b^2 + \frac{M-m}{MNmn} \sum_{i=1}^N S_i^2$$

$$\text{यदि } S_w^2 = \frac{1}{N} \sum_{i=1}^N S_i^2, \quad (28.8)$$

$$\text{बौर } S_u^2 = S_b^2 - \frac{S_w^2}{M} \text{ तो}$$

$$V(\hat{X}) = \frac{N-n}{Nn} S_u^2 + \frac{MN-mn}{MNmn} S_w^2 \quad (28.9)$$

$$\text{यदि } s_u^2 = \frac{1}{n} \sum_{i=1}^n s_i^2 \text{ तथा } s_u^2 + \frac{1}{m} s_w^2 = s_b^2$$

$$= \frac{1}{n-1} \sum_{i=1}^n \left(\bar{x}_i - \frac{\sum x_j}{n} \right)^2 \text{ तो यह आप आभासी से सिद्ध कर सकते हैं कि}$$

$$V(\hat{X}) = \frac{N-n}{Nn} s_u^2 + \frac{MN-mn}{MNmn} s_w^2 \quad (28.10)$$

§ २८८ अनुकूलतम वितरण

यदि हम सब चुनी हुई प्रथम चरणी इकाइयो में से बराबर संख्या में द्वितीय-चरणी इकाइयो का चयन करना चाहें तो हम यह जानना चाहेंगे कि कुल व्यय के दिये होने पर कितनी प्रथम चरणी इकाइयाँ और प्रत्येक प्रथम चरणी इकाई में से कितनी द्वितीय चरणी इकाइयो का चयन किया जाय ।

हम निम्नलिखित व्यय फलन का उपयोग करेंगे

$$C = a + bn + dmn$$

जहाँ a कुछ ऐसा व्यय है जिसका प्रतिद्वंद्व परिमाण से कुछ संबंध नहीं है, b प्रत्येक प्रथम-चरणी इकाई से संबंधित बौर d प्रत्येक द्वितीय चरणी इकाई से संबंधित व्यय है ।

इसी प्रकार प्रसरण फलन को निम्नलिखित रूप में रखा जा सकता है

$$V = -\frac{1}{N} S_b^2 + \frac{1}{n} \left[S_b^2 - \frac{\sum_{i=1}^N M_i S_i^2}{NM^2} \right] + \frac{1}{mn} \frac{\sum_{i=1}^N M_i^2 S_i^2}{NM^2}$$

$$= a + \frac{b}{n} + \frac{d}{mn}$$

कुल व्यय C_0 के दिये होने पर ह्य m और n के ऐसे मानों का पता चलाना चाहते हैं जो प्रसरण को निम्नतम कर दें। इसके लिए हम एक परिमाण Q की परिभाषा देते हैं।

$$Q = a + \frac{b}{n} + \frac{d}{mn} + \lambda [a + bn + dmn - C_0]$$

m और n को प्राप्त करने के लिए निम्नलिखित समीकरण हैं

$$(i) \quad \frac{\partial Q}{\partial m} = 0 \quad \text{अथवा} \quad \frac{d}{m^2 n} = \lambda d n$$

$$\text{अथवा } mn = \frac{1}{\sqrt{\lambda}} \sqrt{\frac{d}{d}} \quad (28 \text{ I1})$$

$$(ii) \quad \frac{\partial Q}{\partial n} = 0$$

$$\text{अथवा } \frac{b'}{n^2} + \frac{d}{mn^2} = \lambda [b + dm]$$

$$\text{अथवा } \frac{b}{n} + \frac{d}{mn} = \lambda [bn + dmn]$$

$$\text{अथवा } \frac{b}{n} = \lambda bn$$

$$\text{अथवा } n = \frac{1}{\sqrt{\lambda}} \sqrt{\frac{b}{b}} \quad (28 \text{ I2})$$

समीकरण (28.11) को (28.12) से विभाजित करने पर

$$m = \sqrt{\frac{d'/d}{b'/b}}$$

इस प्रकार यह प्रतीत होता है कि यदि व्यय-फलन उपरिलिखित है तो m का अनुकूलतम मान कुल व्यय से स्वतंत्र है। कुल व्यय के विभिन्न मान दिये होने पर केवल n के मान में अंतर आयेगा और m का मान स्थिर रहेगा।

यह स्पष्ट है कि a, b, d तथा a', b', d' हमें पहिले से ज्ञात नहीं हो सकने। इन प्राचलों के मान मालूम करने के लिए छोटे पैमाने पर एक आरंभिक सर्वेक्षण की आवश्यकता होती है। इसके आधार पर इन प्राचलों का प्राक्कलन किया जाता है।

§ २८.९ उदाहरण

समष्टि में कुल 20,000 प्रथम-चरणी इकाइयाँ थी जिनमें से प्रारंभिक सर्वेक्षण में 20 चुनी गयी। प्रत्येक प्रथम-चरणी इकाई में 1,000 द्वितीय-चरणी इकाइयाँ थी। चुनी हुई प्रथम-चरणी इकाइयों में से प्रत्येक में से 3 द्वितीय-चरणी इकाइयाँ चुनी गयी। इस प्रकार हमें निम्नलिखित सामग्री प्राप्त हुई

$$s_w^2 = \frac{1}{20} \sum_{i=1}^{20} \frac{3}{2} \frac{(x_{ij} - \bar{x}_i)^2}{2} = 12.24$$

$$s_b^2 = \frac{1}{19} \sum_{i=1}^{20} \left(\bar{x}_i - \frac{1}{20} \sum_{i=1}^{20} \bar{x}_i \right)^2 = 25.13$$

$$\therefore s_u^2 = s_b^2 - \frac{1}{3} s_w^2 \approx 21.05$$

\therefore प्रसरण फलन का निम्नलिखित प्राक्कलन होगा

$$V = \left(\frac{1}{n} - \frac{1}{20,000} \right) \times 21.05 + \left(\frac{1}{mn} - \frac{1}{20,000 \times 1,000} \right) \times 12.24$$

$$= \frac{21.05}{n} + \frac{12.24}{mn}$$

$$\therefore a' = 0, b' = 21.05, d' = 12.24$$

इसके अलावा हमें निम्नलिखित a , b और c के मान प्राप्त हुए।

$$a = 1,000 \text{ रुपए}, b = 42 \text{ 10 रुपए}, d = 6 \text{ 12 रुपए}$$

$$\begin{aligned} m &= \sqrt{\frac{42 \text{ 10} \times 12 \text{ 24}}{21.05 \times 6.12}} \\ &= 2 \end{aligned}$$

यदि सर्वेक्षण के लिए कुल 5,000 रुपए मजूर हुए हों तो

$$5,000 \text{ रुपए} = 1,000 \text{ रुपए} + (42 \text{ 10}) n \text{ रुपए} + (6 \text{ 12}) mn \text{ रुपए}$$

$$\text{परंतु } mn = 2$$

$$\therefore n = \frac{5,000 - 1,000}{42 \text{ 10} + 12 \text{ 24}}$$

$$\therefore = \frac{4,000}{54 \text{ 34}}$$

$$= 71$$

अध्याय २९

सामूहिक प्रतिचयन (Cluster Sampling)

§ २९.१ सामूहिक प्रतिचयन

यदि हमें m परिमाण का एक प्रतिदर्श चुनना हो तो समष्टि को m इकाइयों के समूहों में विभाजित करके इनमें से एक समूह को चुना जा सकता है। इस प्रकार के प्रतिचयन को सामूहिक प्रतिचयन कहते हैं। यह आवश्यक नहीं है कि प्रत्येक समूह में इकाइयों की संख्या बराबर ही हो अथवा केवल एक ही समूह का चयन किया जाय। उदाहरण के लिए किसान परिवारों के सर्वेक्षण में यदि हम कुछ गाँवों को चुने और इन गाँवों के सभी किसान परिवारों का सर्वेक्षण करें तो यह एक सामूहिक प्रतिचयन होगा। आप सामूहिक प्रतिचयन को द्वि-वर्णी प्रचयन का एक सीमांत रूप समझ सकते हैं जिसमें $m_1 = M$

मान लीजिए कुल समष्टि को K समूहों में विभाजित किया गया है और इसमें से k समूहों का सरल यादृच्छिक प्रतिचयन किया गया है। i -वें चुने हुए समूह के लिए गुण X के योग को X_i से सूचित किया जायगा।

$$E \left[\frac{K}{k} \sum_{i=1}^k x_i \right] = \sum_{i=1}^K X_i = X \quad \dots\dots(29.1)$$

इस प्रकार इस प्रचयन-विधि के लिए गुण-समष्टि-योग का प्राक्कलन

$$\hat{X} = \frac{K}{k} \sum_{i=1}^k x_i \text{ है।}$$

§ २९.२ अनुपाती प्राक्कलन

यदि हमें समष्टि की कुल इकाइयों की संख्या $M = \sum_{i=1}^k M_i$ ज्ञात हो तो हम X के इस प्राक्कलन को M से भाग देकर $\bar{X} = \frac{X}{M}$ का प्राक्कलन प्राप्त कर सकते हैं। परंतु बहुधा हमें समष्टि की कुल इकाइयों की संख्या ज्ञात नहीं होनी। यदि हम प्रति किसान परिवार आय का प्राक्कलन करना चाहें तो हमें कुल किसान परिवारों की संख्या ज्ञात

होनी चाहिए, तभी हम इस प्रकार के प्राक्कलन का प्रयोग कर सकते हैं। जिस प्रकार किसान परिवारों की कुल आय का प्राक्कलन किया गया है उसी प्रकार कुल किसान परिवारों की सख्या का भी प्राक्कलन किया जा सकता है। इन दो प्राक्कलों के अनुपात से हमें प्रति किसान परिवार आय का एक प्राक्कलन प्राप्त हो जाता है। यदि i -वें चुने हुए गाँव में किसान परिवारों की सख्या m_i हो तो कुल परिवार सख्या का प्राक्कलक

$$\hat{M} = \frac{K}{k} \sum_{i=1}^k m_i \quad \dots (29.2)$$

$$\therefore \frac{\hat{X}}{\hat{M}} = \frac{\sum_{i=1}^k x_i}{\sum_{i=1}^k m_i} \quad \dots (29.3)$$

इस प्रकार की प्राक्कलन विधि को अनुपाती-प्राक्कलन (ratio estimation) कहते हैं क्योंकि यह दो प्राक्कलों के अनुपात से प्राप्त होता है। यह प्राक्कलन अभिनत नहीं होता। यदि M का ज्ञान हो तो दो प्रकार के प्राक्कलक हो सकते हैं।

$$(1) \quad \hat{X}_2 = \frac{\hat{X}}{\hat{M}}$$

$$(2) \quad \hat{X}_1 = \frac{\hat{X}}{\hat{M}}$$

यदि विभिन्न गाँवों की प्रति किसान-परिवार-आय में विशेष अंतर न हो परंतु किसान परिवारों की सख्या में बहुत अंतर हो तो यह देखा जा सकता है कि दूसरा प्राक्कलक \hat{X}_2 अभिनत होते हुए भी \hat{X}_1 से उत्तम होगा।

§ २९.३ व्यवस्थित-प्रतिचयन (Systematic Sampling)

सामूहिक प्रतिचयन का एक विशेष रूप व्यवस्थित-प्रतिचयन है। मान लीजिए कि समष्टि में कुल nk इकाइयाँ हैं जिनमें से n इकाइयों का एक प्रतिदर्श चुनना है। यदि n बहुत बड़ी सख्या हो तो इस परिमाण के सरल यादृच्छिक प्रतिचयन में काफी समय लग सकता है। इससे अधिक सरल विधि निम्नलिखित है।

सरल यादृच्छिक प्रतिचयन द्वारा 1 से k के बीच में कोई सख्या चुन लीजिए। मान लीजिए यह सख्या r है। यदि i -वीं इकाई को U_i से सूचित किया जाय तो प्रतिदर्श प्राप्त करने के लिए निम्नलिखित इकाइयाँ चुन लीजिए —

$$U_r, U_{r+k}, \dots, U_{r+2k}, \dots, U_{r+ik}, \dots, U_{r+(n-1)k}$$

इस प्रकार के प्रतिचयन को व्यवस्थित प्रतिचयन, प्रथम चुनी सख्या r को यादृच्छिक आरम्भ (random start) और k को प्रतिचयन अंतराल (sampling interval) कहते हैं।

यह देखा जा सकता है कि यह भी सामूहिक प्रतिचयन ही का एक विशेष रूप है। इसमें समष्टि को n इकाइयों के निम्नलिखित k समूहों में विभाजित किया जाता है।

$$U_r, U_{r+k}, U_{r+2k}, \dots, U_{r+(n-1)k}; r=1, 2, 3, \dots, k$$

व्यवस्थित प्रतिचयन द्वारा हम इनमें से एक समूह को चुन लेते हैं।

§ २९४ प्रारोहक समूह (Overlapping clusters) बहुधा समष्टि की कुल इकाइयों की सख्या N को प्रतिदर्श परिमाण n और किसी पूर्णांक के गुणन फल के रूप में नहीं रखा जा सकता। उदाहरण के लिए यदि 107 इकाइयों में से 10 को चुनना हो तो ऊपर लिखी विधि नहीं अपनायी जा सकती। इसके लिए जिस विधि का प्रयोग किया जाता है, वह नीचे दी हुई है।

पहले 1 और N के बीच एक सख्या r को यादृच्छिक प्रतिचयन द्वारा चुना जाता है। यदि $\frac{N}{n} = k \frac{l}{n}$ (अर्थात् n का भाग N में k बार जाता है और l शेष बच जाता है, दूसरे शब्दों में k उन सब पूर्ण सख्याओं में से महत्तम है जो $\frac{N}{n}$ से छोटी हैं) तो इस चयन में r को यादृच्छिक आरम्भ और k को अंतराल लिया जाता है। इस प्रकार चुना हुआ प्रतिदर्श निम्नलिखित होता है

$$U_r, U_{r+k}, U_{r+2k}, \dots, U_{r+ik}, \dots, U_{r+(n-1)k}$$

यहाँ जब $r+ik > N$ हो जाए तब U_{r+ik} के स्थान में U_{r+ik-N} चुना जाता है। उदाहरण के लिए यदि $N=107$, $n=10$ तो $k=10$ । यदि 1 और 107 के बीच चुनी हुई सख्या 89 हो तो प्रतिदर्श निम्नलिखित होगा

$$U_{89}, U_{99}, U_{109}, U_{119}, U_{129}, U_{139}, U_{149}, U_{159}, U_{169}, U_{179},$$

$$यानी U_{89}, U_{99}, U_{9}, U_{19}, U_{29}, U_{39}, U_{49}, U_{59}, U_{69}, U_{79}$$

इस प्रकार के प्रतिचयन को भी व्यवस्थित प्रतिचयन कहते हैं परन्तु जिन समूहों को चुना जा सकता है वे परस्पर अपवर्जी (exclusive) नहीं होते बल्कि प्रारोहक

(overlapping) होते हैं। इस प्रकार के व्यवस्थित प्रतिचयन के लिए भी प्रतिदर्श-माध्य समष्टि-माध्य वा अनिभिनत प्राक्कलन होता है।

§ २९५ सामूहिक प्रतिचयन में प्रसरण

यह स्पष्ट है कि सामूहिक प्रतिचयन के लिए यदि प्रसरण को V_{σ} से सूचित किया जाय तो

$$V_{\sigma} = \frac{K(K-k)}{k} \times \frac{\sum_{i=1}^K \left(X_i - \frac{\sum_{i=1}^K X_i}{K} \right)^2}{K-1} \quad (294)$$

§ २९६ प्रसरण का प्राक्कलक

$$\hat{V}_{\sigma} = \frac{K(K-k)}{k} \frac{\sum_{i=1}^k \left(v_i - \frac{\sum_{i=1}^k v_i}{k} \right)^2}{k-1} \quad (295)$$

यदि प्रतिदर्श में केवल एक समूह चुना जाय जैसा कि व्यवस्थित प्रतिचयन में होता है तो समष्टि योग के प्राक्कलक के प्रसरण का प्राक्कलन नहीं किया जा सकता।

§ २९७ सामूहिक और सरल यादृच्छिक प्रतिचयन की तुलना

आप यह जानना चाहेंगे कि सरल यादृच्छिक प्रतिचयन की तुलना में सामूहिक प्रतिचयन से प्राप्त प्राक्कलन का प्रसरण किस अवस्था में अधिक और किस अवस्था में कम होता है।

$$\begin{aligned} (N-1)S^2 &= \sum_{i=1}^K \sum_{j=1}^n \left(X_{ij} - \frac{\sum_{i=1}^K \sum_{j=1}^n X_{ij}}{K} \right)^2 \\ &= \sum_{i=1}^K \sum_{j=1}^n \left(X_{ij} - \frac{\sum_{j=1}^n X_{ij}}{n} \right)^2 + \frac{1}{n} \sum_{i=1}^K \left(X_i - \frac{\sum_{j=1}^n X_{ij}}{K} \right)^2 \\ &= (n-1) \sum_{i=1}^K S_i^2 + \frac{1}{n} \sum_{i=1}^K \left(X_i - \frac{\sum_{j=1}^n X_{ij}}{K} \right)^2 \end{aligned}$$

यदि हम $\frac{1}{K} \sum_{i=1}^K S_i^2$ को S_w^2 से सूचित करें तो

$$V_e = \frac{nK(K-k)}{k(K-1)} \left[(nK-1)S^2 - K(n-1)S_w^2 \right] \quad (296)$$

$$\begin{aligned} \text{तथा } V_{ran} &= \frac{Kn^2(K-k)}{nk} S^2 \\ &= \frac{nK(K-k)}{k} S^2 \end{aligned} \quad (297)$$

सरल यादृच्छिक प्रतिचयन से सामूहिक प्रतिचयन उत्तम होगा यदि

$$(nK-1)S^2 - K(n-1)S_w^2 < (K-1)S^2$$

$$\text{अथवा } K(n-1) [S^2 - S_w^2] < 0$$

$$\text{अथवा } S^2 < S_w^2$$

S_w^2 समूहाम्यन्तरिक प्रसरण है। हम देखते हैं कि समूहाम्यन्तरिक प्रसरण कुल समष्टि के प्रसरण से अधिक हो तो सामूहिक प्रतिचयन अधिक उत्तम होता है। यदि विभिन्न समूहों के बनाने की हमें स्वतंत्रता हो और व्यय में इन समूहों के निर्माण से कुछ अंतर न पड़े तो यह निर्माण इस प्रकार करना चाहिए कि वे अधिक-से-अधिक विषमता (heterogenous) हो। अर्थात् समूहाम्यन्तरिक प्रसरण अधिक-से-अधिक हो।

अनुपाती प्राक्कलन (Ratio Estimation)

§ ३०.१ अनुपात का प्राक्कलन

यदि दो समष्टि-योग X और Y के अनुपात $R = \frac{Y}{X}$ का प्राक्कलन करना हो

तो X और Y के अलग-अलग प्राक्कलनो \hat{X} तथा \hat{Y} के अनुपात $\hat{R} = \frac{\hat{Y}}{\hat{X}}$ का

इसके लिए प्रयोग किया जाता है। यह सिद्ध किया जा सकता है कि इस प्रकार का प्राक्कलन अभिनत नहीं होता।

यदि प्रतिदर्श—परिमाण यथेष्ट रूप से बड़ा हो तो इस प्राक्कलक की अभिनति और माध्यवर्ग-त्रुटि (mean square error) का सन्निकटन \hat{Y} और \hat{X} के प्रसरणों और सहप्रसरणों तथा अभिनतियों के फलन के रूप में किया जा सकता है। ये सन्निकटन निम्नलिखित हैं—

§ ३०.२ अनुपाती प्राक्कलक अभिनति

$$\begin{aligned} B(\hat{R}) &= E(\hat{R} - R) \\ &= E\left[\frac{1}{\hat{X}}(\hat{Y} - R\hat{X})\right] \end{aligned}$$

$$\begin{aligned} \text{परन्तु } \frac{1}{\hat{X}} &= \frac{1}{X'} \left(1 + \frac{\hat{X} - X'}{X'}\right) \quad \text{जहाँ } E(\hat{X}) = X' \\ &= \frac{1}{X'} \left[1 - \frac{\hat{X} - X'}{X'}\right] \end{aligned}$$

$$\begin{aligned} \therefore B(\hat{R}) &= \frac{1}{X'} E[\hat{Y} - R\hat{X}] \left(1 - \frac{\hat{X} - X'}{X'}\right) \\ &= \frac{1}{X'} \{ [E(\hat{Y}) - Y] - R [E(\hat{X}) - X] \} \end{aligned}$$

$$\begin{aligned}
 & -\frac{1}{\bar{X}^2} [\text{Cov}(\hat{X}, \hat{Y}) - R V(\hat{X})] \\
 & = \frac{1}{\bar{X}^2} [B(\hat{Y}) - R B(\hat{X})] + \frac{1}{\bar{X}^2} [R V(\hat{X}) - \text{Cov}(\hat{X}, \hat{Y})] \quad (301)
 \end{aligned}$$

जहाँ $B(\hat{Y}), B(\hat{X})$ से हमारा तात्पर्य क्रमशः \hat{Y} और \hat{X} की अभिनतियों (biases) से और $\text{Cov}(\hat{X}, \hat{Y})$ से हमारा तात्पर्य \hat{X} और \hat{Y} के सहप्रसरण से है।

यदि \hat{Y} और \hat{X} क्रमशः Y और X के अनभिनत प्रायकलक हों तो $B(\hat{Y}) = -B(\hat{X}) = 0$ और $X' = X$ । इस दशा में

$$B(\hat{R}) = \frac{1}{\bar{X}^2} [R V(\hat{X}) - \text{Cov}(\hat{X}, \hat{Y})] \quad (302)$$

यदि प्रतिषयन विधि सरल यादृच्छिक हो तो

$$V(\hat{X}) = \frac{N(N-n)}{n} \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1}$$

$$\text{Cov}(\hat{X}, \hat{Y}) = \frac{N(N-n)}{n} \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{N-1}$$

$$\text{तथा } V(\hat{Y}) = \frac{N(N-n)}{n} \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N-1}$$

इसलिए $(\hat{R}) = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}$, और बड़े प्रतिदर्शों के लिए $B(\hat{R})$ का निम्नलिखित

सन्निकटन लिया जा सकता है।

$$\begin{aligned}
 B(\hat{R}) & = \frac{1}{\bar{X}^2} \frac{N(N-n)}{n(N-1)} \left[R \left\{ \sum_{i=1}^N X_i^2 - N\bar{X}^2 \right\} \right. \\
 & \quad \left. - \left\{ \sum_{i=1}^N X_i Y_i - N\bar{X}\bar{Y} \right\} \right] \\
 & = \frac{1}{\bar{X}^2} \frac{N(N-n)}{n(N-1)} \sum_{i=1}^N X_i (R X_i - Y_i) \quad (303)
 \end{aligned}$$

§ ३०.३ अभिनति का प्राक्कलन :

$$\hat{B}(\hat{R}) = \frac{1}{\hat{X}^2} \frac{N(N-n)}{n(n-1)} \sum_{i=1}^n x_i (\hat{R}x_i - Y_i) \quad \dots (30.4)$$

§ ३०.४ अनुपाती प्राक्कलन की माध्य-वर्ग-त्रुटि

यदि प्रतिदर्श परिमाण इतना बड़ा हो कि \hat{X} और X' में विक्षेप अंतर न हो तो \hat{R} की माध्य-वर्ग-त्रुटि (M.S.E) होगी

$$\begin{aligned} M S E (\hat{R}) &= E(\hat{R}-R)^2 \\ &= E \frac{1}{\hat{X}^2} (\hat{Y}-R\hat{X})^2 \\ &= \frac{1}{\hat{X}^2} E (\hat{Y}-R\hat{X})^2 \\ &= \frac{1}{\hat{X}^2} E[(\hat{Y}-Y)-R(\hat{X}-X)]^2 \\ &= \frac{1}{\hat{X}^2} [M S E (\hat{Y}) - 2R M P E(\hat{X}, \hat{Y}) \\ &\quad + R^2 M S E (\hat{X})] \quad \dots (30.5) \end{aligned}$$

जहाँ $M P E (\hat{X}, \hat{Y}) = E(\hat{X}-X)(\hat{Y}-Y)$

यदि प्रक्षयन सरल यादृच्छिक हो तो

$$M S E (\hat{R}) = \frac{1}{\hat{X}^2} \frac{N(N-n)}{n(N-1)} \sum_{i=1}^n (Y_i - R X_i)^2 \quad \dots (30.6)$$

ऊपर दिये $M S E (\hat{R})$ के समिकटन का प्राक्कलन नीचे दिए हुये सूत्र द्वारा किया जा सकता है।

$$M \hat{S} E. (\hat{R}) = \frac{1}{\hat{X}^2} \frac{N(N-n)}{n(N-1)} \sum_{i=1}^n (Y_i - \hat{R} X_i)^2 \quad \dots (30.7)$$

§ ३०.५ समष्टि-योग का अनुपाती-प्राक्कलन

बहुधा समष्टि की प्रत्येक इकाई के लिए किसी गुण x का मान ज्ञात होता है। यदि एक प्रतिदर्श के आधार पर $R = \frac{Y}{X}$ का अनुपाती प्राक्कलन किया जाय तो इस प्राक्कलन को X से गुणा करने पर हमें एक प्राक्कलन Y का प्राप्त होता है जो \hat{Y} से भिन्न है। इस प्रकार से प्राप्त प्राक्कलक को हम Y_{ra} से सूचित करेंगे।

§ ३०६ अनुपाती-प्राक्कलन और साधारण अनभिन्नत प्राक्कलन की तुलना -

$$\begin{aligned} & V(\hat{Y}) - MS E(\hat{Y}_{rat}) \\ &= V(\hat{Y}) - [V(\hat{Y}) - 2RCov(\hat{X}, \hat{Y}) + R^2V(\hat{X})] \\ &= Y^2 \left[\frac{2Cov(\hat{X}, \hat{Y})}{XY} - \frac{V(\hat{X})}{X^2} \right] \end{aligned}$$

∴ \hat{Y}_{rat} , \hat{Y} से अधिक उत्तम है यदि

$$\frac{Cov(\hat{X}, \hat{Y})}{XY} > \frac{1}{2} \frac{V(\hat{X})}{X^2}$$

यदि $Cov(\hat{X}, \hat{Y}) = \rho_{\hat{X}\hat{Y}} \sigma_{\hat{X}} \sigma_{\hat{Y}}$

तथा $V(\hat{X}) = \sigma_{\hat{X}}^2$

तो \hat{Y}_{rat} , \hat{Y} से उत्तम होगा यदि

$$\rho_{\hat{X}\hat{Y}} \frac{\sigma_{\hat{X}}}{X} \frac{\sigma_{\hat{Y}}}{Y} > \frac{1}{2} \left(\frac{\sigma_{\hat{X}}}{X} \right)^2$$

अथवा $\rho_{\hat{X}\hat{Y}} > \frac{1}{2} \frac{(\sigma_{\hat{X}}/X)}{(\sigma_{\hat{Y}}/Y)} = \frac{1}{2} \frac{CV(\hat{X})}{CV(\hat{Y})}$

यहाँ $CV(\hat{X})$ तथा $CV(\hat{Y})$ से हमारा तात्पर्य क्रमशः \hat{X} और \hat{Y} के विचरण-गुणाको (coefficients of variation) से है।

$$CV(\hat{X}) = \frac{\sigma_{\hat{X}}}{X} \text{ तथा } CV(\hat{Y}) = \frac{\sigma_{\hat{Y}}}{Y}$$

बहुधा जिस प्रकार की स्थिति में अनुपात का उपयोग किया जाता है उसमें आशा की जाती है कि $CV(\hat{X})$ और $CV(\hat{Y})$ प्रायः बराबर होंगे। इसलिए यदि $\rho_{\hat{X}\hat{Y}}$ का मान $\frac{1}{2}$ से अधिक हो तो हम \hat{Y}_{rat} उपयोग को अधिक उपयुक्त समझेंगे। इसके अतिरिक्त यदि प्रत्येक इकाई के लिए $Y_i = RX_i$, तो $\hat{Y}_{rat} = Y$ और \hat{Y}_{rat} अनभिन्नत तथा यथार्थ होता है। परन्तु साधारणतया ऐसी स्थिति नहीं पायी जाती। यदि Y_i और X_i के अनुपात में विशेष विचलन न हो तो आशा की जा सकती है कि \hat{Y}_{rat} की त्रुटि बहुत कम होगी। इसलिए इस प्रकार की स्थिति में \hat{Y} के स्थान पर \hat{Y}_{rat} का उपयोग अधिक उपयुक्त होगा।

§ ३०७ उदाहरण :—

1951 में जिला हमीरपुर की कुल जनसंख्या 590,731 थी। 1958 में जनसंख्या का प्राक्कलन करने के लिए जिले के 911 गावा में से 20 का सरल यादृच्छिक प्रतिनयन किया गया। इस प्रतिदर्श के लिए

$$\sum_{i=1}^{20} y_i = 27,443$$

$$\sum_{i=1}^{20} y_i^2 = 96,304,953$$

$$\sum_{i=1}^{20} x_i = 24,698$$

$$\sum_{i=1}^{20} x_i^2 = 75,779,814$$

$$\sum_{i=1}^{20} x_i y_i = 85,289,177$$

$$\therefore \hat{R} = \frac{\sum_{i=1}^{20} y_i}{\sum_{i=1}^{20} x_i} = 1.11114$$

$$\hat{Y} = \frac{911}{20} \times 27,443$$

$$= 1,250,029$$

$$\hat{Y}_{rat} = 590,731 \times 1.11114$$

$$= 656,385$$

$$V(\hat{Y}) = \left[96,304,953 - 30,795,162 \right] \frac{911 \times 891}{20 \times 19}$$

$$= 65,509,791 \times \frac{8,11,701}{380}$$

$$MSE(\hat{Y}_{rat}) = \sum_{i=1}^{20} y_i^2 - 2\hat{R} \left[\sum_{i=1}^{20} x_i y_i + \hat{R}^2 \sum_{i=1}^{20} x_i^2 \right] \times \frac{811,701}{380}$$

$$= 328,704 \times \frac{811,701}{380}$$

क्योंकि $V(\hat{Y})$ का मान $MSE(\hat{Y}_{rat})$ से लगभग 20 गुना है इसलिए यह स्पष्ट है कि अनुपाती प्राक्कलन \hat{Y}_{rat} साधारण प्राक्कलन \hat{Y} से उत्तम है।

§ ३०८ प्रतिदर्श-परिमाण

यह ध्यान देने योग्य बात है कि ऊपर दिये हुए अभिनति और प्रसरण के सूत्र केवल सन्निकटन हैं जो प्रतिदर्श परिमाण के मयेष्ट रूप से बड़े होने पर ही उपयुक्त समझे जा

सकते हैं। कितने बड़े प्रतिदर्श को यथेष्ट रूप से बड़ा मानना चाहिए यह ठीक से नहीं कहा जा सकता। विभिन्न समष्टियों के लिए विभिन्न संख्याएँ यथेष्ट हैं। यह इस पर निर्भर करता है कि X_i और Y_i का अनुपात कहां तक अचर है। साधारणतया यदि प्रतिदर्श परिमाण 30 से अधिक हो और इतना हो कि $CV(\bar{X})$ तथा $CV(\bar{Y})$ दोनों ही १० प्रतिशत से कम हों तो इसको काफी बड़ा समझा जा सकता है।

सारणी संख्या 301

1951 और 1958 में जिला हमीरपुर के कुछ गांवों की जनसंख्या

ग्राम संख्या	1951 की जन संख्या	1958 की जन संख्या	अनुपात
i	X_i	Y_i	Y_i/X_i
(1)	(2)	(3)	(4)
1	1,865	1,905	
2	368	399	
3	817	1,025	
4	1,627	2,003	
5	651	726	
(6)	270	238	0.8667
7	1,644	1,712	
8	564	590	
9	488	480	
(10)	6,942	8,042	1.1585
11	792	980	
12	2,121	2,222	
13	222	290	
14	736	872	
(15)	563	614	1.0906
16	165	177	
(17)	1,091	1,201	1.1008
18	3,026	3,117	
19	469	521	
20	277	329	
कुल	24,698	27,443	

विभिन्न-प्रायिकता प्रचयन (Selection with Varying Probabilities)

§ ३११ चयन विधि

अभी तक हमने जितनी भी प्रतिचयन विधियों का अध्ययन किया है वे एक या अधिक स्तरों में, एक या अधिक चरणों में, इकाइयों अथवा समूहों का सरल यादृच्छिक प्रतिचयन ही थीं। परंतु हम अन्य प्रकार से इकाइयों को चुनने की भी कल्पना कर सकते हैं जिसमें यद्यपि चयन की प्रायिकता का प्रत्येक प्रतिदश के लिए परिकलन किया जा सकता हो परंतु ये प्रायिकताएँ सब प्रतिदशों के लिए बराबर न हों। इस प्रकार की प्रतिचयन विधि को विभिन्न प्रायिकता चयन (selection with varying probabilities) कहते हैं।

मान लीजिए कि कुल इकाइयों की संख्या N है। इनको हम $U_1, U_2, \dots, U_n, \dots, U_N$ से सूचित करेंगे। हम पहिले से निश्चित कर सकते हैं कि इन इकाइयों के प्रतिचयन की प्रायिकता क्रमशः $p_1, p_2, \dots, p_n, \dots, p_N$ होगी। इसमें हमें इकाइयों के द्वारा चुने जाने पर प्रतिबन्ध लगाने की कोई आवश्यकता नहीं है। मान लीजिए P एक ऐसी पूर्ण संख्या है जिससे गुणा करने पर ये सब प्रायिकताएँ पूर्ण संख्याओं में परिणत हो जाती हैं। यदि P और इन प्रायिकताओं के गुणनफल को क्रमशः $P_1, P_2, \dots, P_i, \dots, P_N$ से सूचित किया जाय तो निम्नलिखित चयन विधि से हम इच्छित प्रायिकताओं को प्राप्त कर सकते हैं। [यह उसी दशा में संभव है जब सब प्रायिकताएँ परिमेय संख्याएँ (rational numbers) हों।]

हम P_1, P_2, \dots, P_N को क्रम से एक स्तंभ में लिखकर इनके संचयी योग (cumulative totals) को दूसरे स्तंभ में लिख सकते हैं जैसा नीचे की सारणी में दिया हुआ है।

सारणी संख्या 311

क्रम संख्या ;	P^X प्रायिकता $=P_i$	सचयी योग i $\sum_{j=1} P_j = S_i$
(1)	(2)	(3)
1	P_1	$P_1 = S_1$
2	P_2	$P_1 + P_2 = S_2$
3	P_3	$P_1 + P_2 + P_3 = S_3$
⋮		
i	P_i	i $\sum_{j=1} P_j = S_i$
N	P_N	N $\sum_{j=1} P_j = P_N = S_N$

यदि कोई एक संख्या 1 से P तक की संख्याओं में से समान प्रायिकता से चुनी जाय तो उसके S_{i-1} और S_i के बीच में होने की क्या प्रायिकता है? क्योंकि S_{i-1} और S_i के बीच कुल संभव संख्याएँ P_i हैं। इसलिए स्पष्टतया यह प्रायिकता $\frac{P_i}{P} = P_i$ है।

यही वह प्रायिकता है जो हम U_i के चयन के लिए चाहते थे। इसलिए हमारी चयन विधि निम्नलिखित हो सकती है।

1 से P तक की संख्याओं में से एक को समान प्रायिकता से चुन लिया जाय। यह संख्या सारणी में दिये हुए सचयी योगों में से किन्हीं दो (S_{i-1} और S_i) के बीच में पड़ेगी।

इनमें से वह जिससे कम हो अथवा जिसके बराबर हो (अर्थात् S_i) उससे संबंधित ड्रकॉर्ड (U_i) को चुना हुआ माना जायगा।

§ ३१.२ विकल्प विधि

यदि कुल इकाइयों की संख्या बहुत अधिक हो तो ऊपर दिए हुए तरीके से सचयी योगों को प्राप्त करने में बहुत समय और मेहनत लगेगी। इस दशा में एक और विधि है जिसके द्वारा इच्छित प्रायिकताएँ प्राप्त की जा सकती हैं। इस विधि के निम्नलिखित चरण हैं।

(१) १ से N तक की संख्याओं में से किसी एक का समान प्रायिकता से प्रतिचयन किया जाता है। चुनी हुई संख्या को हम r से सूचित करेंगे।

(२) मान लीजिए P' एक ऐसी संख्या है जो किसी भी P_r से कम नहीं है। एक दूसरी संख्या 1 से P' तक की संख्याओं में से समान प्रायिकता से चुनी जाती है। इस चुनी हुई संख्या को r' से सूचित किया जायगा।

(३) यदि $r' \leq P_r$ हो तो हम r वी इकाई U_r को चुन लेते हैं, अन्यथा फिर प्रथम और द्वितीय चरणों को दुहराते हैं जब तक कि हमें इच्छित परिमाण का प्रतिदर्श प्राप्त नहीं हो जाता।

इस विधि द्वारा प्रथम बार में r वी इकाई को चुने जाने की प्रायिकता $\frac{1}{N} \frac{P_r}{P'}$

है। इस घटना की प्रायिकता कि कोई भी इकाई नहीं चुनी जायगी $\left(1 - \frac{P}{NP'}\right)$

है। क्योंकि U_r पहिले, दूसरे, तीसरे इत्यादि प्रयत्नों में चुनी जा सकती है इसलिए इसके चुने जाने की कुल प्रायिकता

$$\begin{aligned}
 P(U_r) &= \frac{1}{N} \frac{P_r}{P'} + \left(1 - \frac{P}{NP'}\right) \frac{1}{N} \frac{P_r}{P'} + \left(1 - \frac{P}{NP'}\right)^2 \frac{1}{N} \frac{P_r}{P'} + \dots \\
 &+ \dots + \left(1 - \frac{P}{NP'}\right)^i \times \frac{1}{N} \frac{P_r}{P'} + \dots + \dots + \dots \\
 &= \frac{1}{N} \frac{P_r}{P'} \sum_{l=0}^{\infty} \left(1 - \frac{P}{NP'}\right)^l \\
 &= \frac{1}{N} \frac{P_r}{P'} \frac{1}{1 - \left(1 - \frac{P}{NP'}\right)} \\
 &= \frac{P_r}{P} = p_r
 \end{aligned}$$

§ ३१.३ प्राक्कलन

यदि चुनी हुई इकाई U_i हो तो $\frac{Y_i}{p_i}$ कुल समष्टि के Y -गुण के योग का एक अनभिन्नत प्राक्कलक है।

$$E\left(\frac{Y_i}{p_i}\right) = \sum_{r=1}^N \frac{Y_r}{p_r} p_r = \sum_{r=1}^N Y_r = Y \dots\dots\dots (31.1)$$

यदि कुल N इकाइया चुनी जायें तो हमें प्रत्येक इकाई से इस प्रकार का एक अनभिन्नत प्राक्कलन प्राप्त हो सकता है। इसलिए इन प्राक्कलकों का माध्य \hat{Y} भी समष्टि योग Y का एक अनभिन्नत प्राक्कलक है।

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{p_i} \dots\dots (31.2)$$

§ ३१.४ प्राक्कलक का प्रसरण

$$\begin{aligned} V(\hat{Y}) &= \frac{1}{n^2} \sum_{i=1}^n V\left(\frac{Y_i}{p_i}\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{r=1}^N \left(\frac{Y_r}{p_r} - Y\right)^2 p_r \\ &= \frac{1}{n} \left[\sum_{r=1}^N \frac{Y_r^2}{p_r} - Y^2 \right] \dots\dots (31.3) \end{aligned}$$

यदि $p_i = k \gamma_i$ $i=1, 2 \dots N$

तो $1 = \sum_{i=1}^N p_i = k \sum_{i=1}^N \gamma_i = k Y$

$$\therefore k = \frac{1}{Y}$$

इस दशा में

$$V(\hat{Y}) = \frac{1}{n} \left[\sum_{r=1}^N \frac{Y_r^2}{Y_r/Y} - Y^2 \right] = 0$$

§ ३१.५ मापानुपाती प्रायिकता

इससे यह पता चलता है कि यदि इकाइया के चयन की प्रायिकताएँ उनके माना के अनुपात में हों तो हमें इस प्रकार के प्राक्कलन से समष्टि योग का अनुमान बिना किसी भ्रुटि के हो जायगा। वास्तव में हम इसकी आशा नहीं कर सकते परन्तु यह सम्भव है कि Y_i और p_i का अनुपात प्रायः अचल हो। इस स्थिति में विभिन्न प्रायिकता चयन बहुत लाभदायक सिद्ध हो सकता है। यदि एक छोटे से सर्वेक्षण द्वारा हम Y_1, Y_2, \dots, Y_N का अनुमान लगा लें और इन अनुमानों को X_1, X_2, \dots, X_N से सूचित करें तो p_i को X_i के अनुपात में लेने से यह आशा की जा सकती है कि Y_i और p_i का अनुपात प्रायः अचल होगा। इसी प्रकार यदि हमें 1958 में प्रत्येक गाँव में फसल का मात्रा ज्ञात है तो 1959 में कुल जिले में फसल के प्राक्कलन के लिए गाँवों के चुनने की प्रायिकताएँ 1958 की पैदावार के अनुपात में ली जा सकती हैं। तात्पर्य यह है कि यदि हम प्रायिकताओं को किसी ऐसे गुण y के अनुपात में लें जिनसे y का अनुपात प्रायः अचल रहने की आशा की जाती है तो यह प्राक्कलन अच्छे फल दे सकता है। किसी इकाई के x मान को इकाई का माप कहा जा सकता है। इस माप के अनुपात में प्रायिकता चयन को मापानुपाती प्रायिकता चयन (*selection with probabilities proportional to size*) कहा जाता है। यदि इस प्राक्कलन को \hat{Y}_{pps} से सूचित किया जाय तो

$$\begin{aligned}\hat{Y}_{pps} &= \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{p_i} \\ &= \frac{X}{n} \sum_{i=1}^n \frac{Y_i}{X_i}\end{aligned}\quad (31.4)$$

$$\text{जहाँ } X = \sum_{i=1}^N X_i \quad (31.5)$$

§ ३१.६ प्राक्कलक के प्रसरण का प्राक्कलन

हम जानते हैं कि यदि एक समष्टि का प्रसरण σ^2 हो और उसमें से n परिमाण का एक प्रतिदर्श समान प्रायिकता द्वारा चुना जाय (जिसमें इकाइया के द्वारा चुने

जाने पर कोई रोव न हो) तो s^2 का एक अनभिन्नत प्राक्कलक $s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$

है जहाँ i -वीं चुनी हुई इकाई का मान y_i है। यदि हम $\frac{Y_i}{p_i}$ की समष्टि के प्रसरण का प्राक्कलन करना चाहें तो प्राक्कलक निम्नलिखित होगा।

$$\hat{V}\left(\frac{Y_i}{p_i}\right) = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{Y_i}{p_i} - \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{p_i}\right)^2 \quad (31.6)$$

परंतु हमारे प्राक्कलक का प्रसरण $\frac{Y_i}{p_i}$ के प्रसरण का n वा भाग है इसलिए उसका अनभिन्नत प्राक्कलक निम्नलिखित है

$$\hat{V}(\hat{Y}) = \frac{1}{n(n-1)} \sum_{i=1}^n \left(\frac{Y_i}{p_i} - \hat{Y}\right)^2 \quad (31.7)$$

इकाइयों के माप X के रूप में प्राक्कलक निम्नलिखित होगा

$$\hat{V}(\hat{Y}_{pp}) = \frac{X^2}{n(n-1)} \left[\sum_{i=1}^n \left(\frac{Y_i}{X_i}\right)^2 - \frac{1}{n} \left\{ \sum_{i=1}^n \left(\frac{Y_i}{X_i}\right) \right\}^2 \right] \quad (31.8)$$

§ ३१.७ उदाहरण

सारणी 30 I में एक छोटी-सी समष्टि के लिए उसके माप X और गुण Y के मान दिये हुए हैं। उदाहरण द्वारा समझाया जायगा कि इस माप के अनुपात में प्रायिकता लेकर इकाइयों को किस प्रकार चुना जा सकता है। एक चुने हुए प्रतिदर्श से Y के समष्टि-योग का प्राक्कलन किया जायगा और प्राक्कलक के प्रसरण का प्राक्कलन भी किया जायगा।

हमें समष्टि में से पाँच इकाइयाँ चुननी हैं। सारणी सख्या 31.2 के स्तभ (3)

से हमें पता चलता है कि $X = \sum_{i=1}^{20} X_i = 24,698$ । अब हम पाँच सख्याएँ 1 और

24,698 के बीच में से चुनते हैं जो स्तभ (4) में दी हुई हैं। ये सख्याएँ उन्हीं इकाइयों के सामने लिखी गयी हैं जो इनके द्वारा चुनी हुई हैं। उदाहरण के लिए 5,413 पाँचवें

सारणी संख्या 312

क्रम संख्या i	इकाई का माप X_i	संचयी योग $S_i = \sum_{j=1}^i X_j$	यादृच्छिक संख्या r
(1)	(2)	(3)	(4)
1	1,865	1,865	
2	368	2,233	
3	817	3,050	
4	1,627	4,677	
5	651	5,328	
6	270	5,598	5,413
7	1,644	7,242	
8	564	7,806	
9	488	8,294	
10	6,942	15,236	10,541; 14,608
11	792	16,028	
12	2,121	18,149	
13	222	18,371	
14	736	19,107	
15	563	19,670	19,651
16	165	19,835	
17	1,091	20,926	20,734
18	3,026	23,952	
19	469	24,421	
20	277	24,698	

और छोटे संचयी योगों के बीच की संख्या है इसलिए इसके द्वारा छोटी इकाई को चुना जायगा। इस प्रतिदर्श में छोटी, दसवी, पन्द्रहवी और सत्रहवी इकाई चुनी गयी है। दसवी इकाई द्वारा चुनी गयी है। सारणी संख्या 301 में इन चुनी हुई इकाइयों के लिए Y_i और X_i का अनुपात स्तम्भ (4) में दिया हुआ है।

$$\begin{aligned}\hat{Y}_{PPS} &= \frac{24,698}{5} [0.8667 + 1.1585 + 1.1585 + 1.0906 + 1.1008] \\ &= 24,698 \times 1.0750 \\ &= 27,550\end{aligned}$$

सारणी सख्या 30 I से पता चलता है कि $Y=27,443$ । इस प्रकार \hat{Y}_{PPS} एक बहुत ही अच्छा प्रायिकतन है। आप अन्य प्रतिदग लेबर इसकी और \hat{Y}_{RAS} की तुलना कर सकते हैं।

$$\begin{aligned}\hat{V}(Y_{PPS}) &= \frac{(24698)^2}{5 \times 4} [(0.8667)^2 + 2 \times (1.1585)^2 + (1.0906)^2 \\ &\quad + (1.1008)^2 - \frac{1}{5} (5.3751)^2] \\ &= \frac{(24,698)^2}{5 \times 4} \times 0.05845739 \\ &= 1,782,924\end{aligned}$$

पारिभाषिक-शब्दावली

हिन्दी-अंग्रेजी

अनभिनत—unbiased	आयताकार वटन—rectangular distribution
अनभिनतता—unbiasedness	आसन्न सौष्ठव—goodness of fit
अनभिनत प्राक्कलन—unbiased estimator	इकाई—unit
अनुपानी प्राक्कलन—ratio estimation	उत्क्षेपण—loss
अनन्त अनुक्रम—infinitesimal sequence	उपचार—treatment
अनन्त श्रृंखला—infinitesimal series	उपपत्ति—proof
अपवर्गी—exclusive	उपादान—factors
अभिकल्पना—design	ऊर्ध्व—vertical
अभिधारणाएँ—postulates	एक प्रतीय फलन—linear function
अभिनत परीक्षण—biased test	एक प्रतीय—linear
अभिनति—bias	एक पार्श्वीय वटन—marginal distribution
अवकल कलन—differential calculus	एक-समान अधिकतम—uniformly most
असतत—discrete	ताम्यमान परीक्षण—powerful test
असतत वटन—discrete distribution	एक समान अनभिनत परीक्षण—uniformly unbiased test
असमान—heterogeneous	एकस्वनी—monotonic
असम्भाव्य—improbable	अंतरचतुर्क-परास—inter-quartile range
असममिति—asymmetry	अंतराल प्राक्कलन—interval estimation
अस्वीकृति प्रदेश—region of rejection	अंतर समूह—between groups
असिद्ध—disprove	अंश—numerator
आगमिक विधि—inductive method	
आदर्शिकरण—idealisation	
आपेक्षिक बारंबारता—relative frequency	

आकडे, न्यास—data	दक्ष प्राक्कलक—efficient estimator
आंशिक समाकुलन—partial confounding	दशमक—decile
बहुदता—kurtosis	दार्शनिक-तत्त्व विद्या—meta-physics
कारण और कार्य—cause and effect	द्वि घाती परवलय—parabola of second degree
कुन्तल कोष्ठक—curled brackets	द्विपद वटन—binomial distribution
कुलक—set	द्वि विमतीय यादृच्छिक चर—two dimensional random variable
केन्द्रीय प्रवृत्ति—central tendency	घनात्मक—positive
कोटि—ordinate	निश्चय—criterion
क्रमचय—permutation	नियंत्रण इकाइयाँ—control units
क्रमागत—consecutive	नियंत्रण चाट—control chart
क्रमिक-साहचर्य का सूचकांक—index of order association	नियंत्रित यादृच्छिकीकरण—restricted randomisation
खंड—factors	निरपेक्ष मान—absolute value
गतिविज्ञान—dynamics	निरमन—elimination
गुण साहचर्य—association of attributes	नि शयी—exhaustive
ग्राह्य—admissible	न्यास—data
घात श्रणी—power series	परत लब्ध प्रायिकता— <i>a posteriori</i> probability
घूर्ण—moment	पर्याप्त प्रतिदशक—sufficient statistic
घूर्ण विधि—method of moments	पर्याप्ति—sufficiency
चिकित्सा विज्ञान—medical science	परस्पर अपवर्जी घटनाएँ—mutually exclusive events
टकन—type	परास—range
ढरी—lot	परिकल्पना—hypothesis
तुल्य—equivalent	परिकल्पना की जाँच—testing of hypothesis
तोरण—ogive	परिधि—circumference
त्रुटियों का वटन—law of errors	परिमित—finite
त्वरण—acceleration	
दंडचित्र—bar diagram	
दक्षता—efficiency	

परिमेय संख्या—rational number	प्रतिश्रुति—guarantee
परीक्षण सामर्थ्य—power of test	प्रथम चतुर्थांश—first quartile
पारस्परिक साहचर्य—mutual association	प्रमेय—theorem
पूर्वत नुहील प्रायिकता— <i>a-priori</i> probability	प्रयोग अभिकल्पना—design of experiment
पोषण-संबन्धी शोध-प्रयोग—nutritional research	प्रयोजित गणित—applied mathematics
पोषकता—food value	प्रवृत्तियाँ—tendencies
पक्ति—row	प्रसरण—variance
प्रक्षेप—projection	प्रसरण विदलेपण—analysis of variance
प्रकीर्ण चित्र—scatter diagram	प्रसामान्य—normal
प्रतिचयन अंतराल—sampling interval	प्रसार—dispersion
प्रतिच्छेद—intersection	प्रायकालक—estimator
प्रतिदर्श—sample	प्राचल—parameters
प्रतिदर्शज—statistic	प्राथमिक घटनाएँ—elementary events
प्रतिदर्शज वटन—sampling distribution	प्रायिकता—probability
प्रतिदर्श निरीक्षण योजना—sampling inspection plan	प्रायिकता घनत्व—probability density
प्रतिदर्शी त्रुटि—sampling error	प्रायिकता द्रव्यमान—probability mass
प्रतिबंध—restriction	प्रायिकता वटन—probability distribution
प्रतिबंधी प्रायिकता—conditional probability	प्रायोगिक मूल—experimental error
प्रतिबंधी वटन—conditional distribution	प्रारोहक समूह—overlapping clusters
प्रतिरूप—model	प्रेक्षक—observer
प्रतिशतता बिंदु—percentage points	प्रेक्षणगम्य—observable
प्रतिष्ठा—status	प्रेक्षण त्रुटि—observational error

पवासो वटन—Poisson's distribution	मानकित प्रसामान्य वटन—standardized normal distribution
बहु उपादानिय प्रयोग—factorial experiment	माप—measure
बहुचर—multivariate	मापनी—scale
बहुलक (भूयिष्ठ) —mode	मापानुपाती प्रायिकता—probability proportional to size
बहुलक अंतराल—modal interval	मूल बिंदु—origin
बारवारता—दे० बारवारता	मौसम विज्ञान विभाग—meteorological station
बिंदु प्राक्कलन—point estimation	यथार्थता—precision
बुद्धि परीक्षा—intelligence test	यथार्थ नियम—exact laws
भिन्न—fraction	यादृच्छिक आरम्भ—random start
भुज—abscissa	यादृच्छिक चर—random variable
भुजाक्ष—axis of abscissa	यादृच्छिक प्रयोग—random experiments
मन शारीरिक—psychosomatic	यादृच्छिकीकरण—randomization
महत्तम सभाविता विधि—maximum likelihood method	युगपत् समीकरण—simultaneous equations
मानक—unit	युग्म—pair
माध्य—mean	रूप—shape
माध्य वर्ग आसंग—mean square contingency	लघु गणक—logarithm
माध्य वर्ग विचलन मूल—root mean square deviation	लेखाचित्र—graph
माध्य विचलन—mean-deviation	वक्र आसजन—curve fitting
माध्यांतरिक घूर्ण—moment about the mean	वर्ग—square
माध्यिका—medium	वर्गमूल—square root
मानक विचलन—standard deviation	वर्गित विचलन—squared deviations
मानकित मापनी—standardized scale	वनस्पति प्रजनन—plant breeding
	बारवारता—frequency
	बारवारता बहुभुज—frequency polygon

विकल्प— <i>alternative</i>	संचयी प्रायिकता फलन— <i>distribution function</i>
विचलन— <i>deviation</i>	संतुलित असंपूर्ण ब्लॉक अभिकल्पना— <i>balanced incomplete block design</i>
विभिन्न प्रायिकता चयन — <i>selection with varying probabilities</i>	सपरीक्षण (या प्रयोग विधि)— <i>experimentation</i>
विन्यास— <i>arrangement</i>	समावी— <i>likely</i>
विनिदिष्ट— <i>specify</i>	संयुक्त घटनाएँ— <i>joint events</i>
विश्वास गुणांक— <i>confidence coefficient</i>	संयुक्त वटन— <i>joint distribution</i>
विश्वास अंतराल— <i>confidence interval</i>	संयोज्य— <i>additive</i>
विश्वास्य युक्ति— <i>fiducial argument</i>	संयोज्यता गुण— <i>additive property</i>
विश्वास्य वटन— <i>fiducial distribution</i>	सशय अंतराल— <i>critical region</i>
विषम— <i>odd</i>	सतत— <i>continuous</i>
वेग— <i>velocity</i>	सतत वटन— <i>continuous distribution</i>
वैषम्य— <i>skewness</i> ,	सत्य भासक— <i>plausible</i>
वृष्टि मापक— <i>rain gauge</i>	सन्निकटन— <i>approximation</i>
व्यवस्थित प्रतिचयन ~ <i>systematic sampling</i>	सम— <i>even</i>
व्यास— <i>diameter</i>	सममित— <i>symmetrical</i>
वाततमक— <i>percentile</i>	समष्टि— <i>population (universe)</i>
शिखरता— <i>peakedness</i>	समाकलन— <i>integration</i>
शून्यान्तारिक घूर्ण— <i>raw moment</i>	समाकल— <i>integral</i>
संकेत— <i>notation</i>	समान्तर माध्य— <i>arithmetic mean</i>
संख्यात्मक अभिगणना— <i>arithmetical computations</i>	समानुपाती— <i>proportional</i>
संगत— <i>consistent</i>	समाश्रयण— <i>regression</i>
संगम— <i>union</i>	समाश्रयण गुणांक— <i>regression coefficient</i>
सघटक— <i>component</i>	समाश्रयण रेखा— <i>regression line</i>
संयय— <i>combinations</i>	समाश्रयण वक्र— <i>regression curve</i>
संचयी— <i>cumulative</i>	

समागता परीक्षण—test of homogeneity	सामर्थ्य वक्र—power curve
समूहाम्यतरिक—within group	सामर्थ्यवान्—powerful
समजन—adjustment	सामूहिक प्रतिचयन—cluster sampling
समजित उपचार योग — adjusted treatment total	सारणी—table
सर्वेक्षण—survey	साहचर्य—association
सहकारी घर—concomitant variable	साहचर्य सूचक—index of association
सहज ज्ञान—intuition	सुप्राही—sensitive
सह प्रसरण विहलेपण—analysis of covariance	स्तर—level
सह-संबन्ध—correlation	स्तम्भ—column
सह-संबन्ध गुणांक—correlation coefficient	स्थानांक—coordinate
सहसंबन्धानुपात—correlation ratio	स्थानीयत अभिनत—locally biased
सांख्यिक—statistician	स्थानीयत अधिकतम सामर्थ्यवान्—locally most powerful
सांख्यिकी—statistics	स्वातंत्र्य सख्या—degrees of freedom
सांख्यिकीय नियम—statistical laws	स्वीकृति क्षेत्र—acceptance region
सांख्यिकता स्तर—level of significance	स्वेच्छ—arbitrary
	हर—denominator

अंग्रेजी-हिन्दी

abscissa—भुज	association—साहचर्य
absolute value—निरपेक्ष मान	association of attributes—गुण-साहचर्य
acceleration—त्वरण	asymmetry—असममिति
acceptance region—स्वीकृति क्षेत्र	axis of abscissa—भुजाक्ष
additive—सयोग्य	balanced incomplete block design—मनुकलित असंपूर्ण ब्लॉक अभिकल्पना
additive property—सयोग्यता गुण	bar diagram—दण्डचित्र
adjusted treatment total—सम-जित उपचार योग	between groups sum of square—अंतर समूह वग-योग
adjustment—समजन	bias—अभिन्नति
admissible—प्राह्य	biased test—अभिनत परीक्षण
alternative—विकल्प	binomial distribution—द्विपद वटन
analysis of covariance—सह प्रसरण विश्लेषण	cause and effect—कार्य और कारण
analysis of variance—प्रसरण विश्लेषण	central tendency—केन्द्रीय प्रवृत्ति
a-posteriori probability—परत लक्ष्य प्रायिकता	circumference—परिधि
applied mathematics—प्रयोजित गणित	cluster sampling—समूहिक प्रति-चयन
approximation—सन्निकटन	column—स्तम्भ
a-priori probability—पूर्वत गृहीत प्रायिकता	combination—संघम
arbitrary—स्वेच्छ	component—सघटक
arithmetical computations—सहस्रात्मक अभिगणना	concomitant variable—सहकारी चर
arithmetical mean—समांतर माध्य	conditional distribution—प्रतिबंधी-वटन
arrangement—विन्यास	conditional probability—प्रति-बंधी प्रायिकता

confidence coefficient—विश्वास गुणांक	diameter—व्यास
confidence interval—विश्वास्य अंतराल	differential calculus—अवकल कलन
consecutive—क्रमागत	discrete—असतत
consistency—समति	discrete distribution—असतत वटन
consistent—सगत	dispersion—प्रसार
continuous—मत्त	disprove—असिद्ध
continuous distribution—सतत वटन	distribution function—सन्नयी प्रायिकता फलन
control chart—नियंत्रण चार्ट	dynamics—गति विज्ञान
control units—नियंत्रण इकाइयाँ	efficient estimator—दक्ष प्रायिकलक
coordinate—स्थानांक	efficiency—दक्षता
correlation—सहसंबंध	elementary events—प्राथमिक घटनाएँ
correlation coefficient—सहसंबंध गुणांक	elimination—निरसन
correlation ratio—सहसंबंधानुपात	equivalent—तुल्य
criterion—निकष	estimator—प्रायिकलक
critical region—संशय अंतराल	even—सम
cumulative—संचयी	exact laws—यथार्थ नियम
curled brackets—कुन्तल कोष्ठक	exclusive—अपवर्जी
curve fitting—दृक आसजन	exhaustive—नि शेषी
data—आँकड़े, न्यास	experimental error—प्रायोगिक त्रुटि
decile—दशमक	experimentation—संपरीक्षण, प्रयोग विधि
degrees of freedom—स्वातन्त्र्य सख्या	factorial experiment—बहु-उपा-दानीय प्रयोग
denominator—हर	factor—उपादान, खण्ड
design of experiment—प्रयोग अभिकल्पना	fiducial argument—विश्वास्थ्य युक्ति
deviation—विचलन	fiducial distribution—विश्वास्थ्य वटन
	finite—परिमित
	first kind of error—पहली किस्म की त्रुटि

first quartile—प्रथम चतुर्थक	joint events—संयुक्त घटनाएँ
food value—पोष्टिकता	kurtosis—ककुदना
fraction—भिन्न	law of errors—त्रुटियों का बटन
frequency—बारबारता	level—स्तर
frequency polygon—बारबारता बहुभुज	level of significance—साधकता स्तर
goodness of fit—अगिजन सौष्ठव	likly—संभावनी
graph—रेखा चित्र	linear—एकघातीय
guarantee—प्रतिधुति	linear function—एक घातीय फलन
heterogenous—असमाग	locally biased—स्थानीयत अभिनत
hypothesis—परिकल्पना	locally most powerful—स्थानीयत अधिकतम सामध्यवान्
idealisation—आदर्शीकरण	logarithm—लघुगणक
improbable—असंभाव्य	lot—दरी
index of association—साहचर्य सूचक	main effect—मुख्य प्रभाव
index of order association—क्रमिक साहचर्य का सूचकांक	marginal distribution—एक पार्श्वीय बटन
inductive method—आगमिक विधि	maximum likelihood method—महत्तम संभावितता विधि
infinite sequence—अनंत अनुक्रम	mean—माध्य
infinite series—अनंत श्रणी	mean deviation—माध्य विचलन
integral—समाकलन	mean square contingency—माध्य वग असंग
integration—समाकलन	measure—माप
intelligence test—बुद्धि परीक्षण	median—माध्यिका
inter-quartile range—अतश्चतुर्थक परास	medical science—चिकित्सा विज्ञान
intersection—प्रतिच्छेद	meta-physics—तत्त्वविद्या
interval estimation—अंतराल प्राक्कलन	meteorological station—मौसम विज्ञान विभाग
intuition—सहज ज्ञान	method of moments—पूर्ण विधि
joint distribution—संयुक्त बटन	modal interval—बहुलक अंतराल

mode—बहुलक	peakedness—शिखरता
model—प्रतिरूप	percentage points—प्रतिशतता बिंदु
moment—घूर्ण	percentile—शततमक
moment about the mean— माध्यांतरिक घूर्ण	permutation—क्रमचय
monotonic—एकस्वनी	plant breeding—वनस्पति प्रजनन
multivariate—बहुचर	plausible—सत्य भासक
mutual association—पारस्परिक साहचर्य	point estimation—बिंदु प्राक्कलन
mutual exclusive events—परस्पर अपवर्जी घटनाएँ	Poisson's distribution—प्वासो वटन
normal—प्रसामान्य	population (Universe)—समष्टि
notation—संकेत	positive—धनात्मक
numerator—अद्य	postulate—अभिधारणा
nutritional research—पोषण-सबधी श्वेषणा	power—सामर्थ्य
observable—प्रेक्षण गन्म, प्रेक्ष्य	power curve—सामर्थ्य वक्र
observational error—प्रेक्षण त्रुटि	powerful—सामर्थ्यवान्
observer—प्रेक्षक	power of a test—परीक्षण-सामर्थ्य
odd—विषम	power series—घातश्रेणी
ogive—तोरण	precision—यथार्थता
ordinate—कोटि	probability—प्रायिकता
origin—मूल बिन्दु	probability density—प्रायिकता घनत्व
overlapping clusters—प्रारोहक समूह	probability distribution—प्रायिकता वटन
pair—युग्म	probability mass—प्रायिकता द्रव्य- मान
parabola of second degree—द्वि- घाती परवलय	probability proportional to size—मापानुपाती प्रायिकता
parameter—प्राचल	projection—प्रक्षेप
partial confounding—आंशिक समाकुलन	proof—रूपपत्ति
	proportional—समानुपाती
	psychosomatic—मन शारीरिक

rain gauge—वृष्टि-मापक	sampling error—प्रतिदर्शी त्रुटि
random experiment—यादृच्छिक प्रयोग	sampling inspection plan—प्रतिदर्श निरीक्षण योजना
randomization—यादृच्छिकीकरण	sampling interval—प्रतिचयन अंतराल
random start—यादृच्छिक आरम्भ	scale—मापनी
random variable—यादृच्छिक चर	scatter diagram—प्रकीर्ण चित्र
range—पराम	second kind of error—दूसरी किस्म की त्रुटि
ratio-estimation—अनुपाती प्राक्कलन	selection with varying probabilities—विभिन्न प्रायिकता चयन
rational number—परिमेय संख्या	sensitive—सुग्राही
raw moment—शून्यांतरिक पूर्ण	set—कुलक
real number—वास्तविक संख्या	shape—रूप
rectangular distribution—आयताकार वटन	simultaneous equations—युग्मपत् समीकरण
region of rejection—अस्वीकृति क्षेत्र	skewness—वैषम्य
regression—समाश्रयण	specify—विनिर्दिष्ट
regression coefficient—समाश्रयण गुणांक	square—वर्ग
regression curve—समाश्रयण वक्र	squared deviation—वर्गित विचलन
regression line—समाश्रयण रेखा	square root—वर्गमूल
relative frequency—आपेक्षिक बारवारता	standard deviation—मानक विचलन
restricted randomization—नियंत्रित यादृच्छिकीकरण	standardised normal distribution—मानकित प्रसामान्य वटन
restriction—प्रतिबंध	standardised scale—मानकित मापनी
root mean square deviation—माध्य वर्ग-विचलन मूल	statistical laws—सांख्यिकीय नियम
row—पंक्ति	statistics—सांख्यिकी
sample—प्रतिदर्श	status—प्रतिष्ठा
sampling distribution—प्रतिदर्श वटन	sufficiency—पर्याप्त
	sufficient—पर्याप्त
	sufficient statistic—पर्याप्त प्रतिदर्शज

survey—सर्वेक्षण	unbiased—अनभिन्नत
symmetrical—सममित	unbiased estimator—अनभिन्नत प्राक्कलक
systematic sampling—व्यवस्थित प्रतिचयन	unbiasedness—अनभिन्नता
table—सारणी	uniformly most powerful test एक समान अधिकतम सामर्थ्यवान् परीक्षण
tendency—प्रवृत्ति	uniformly unbiased test—एक- समान अनभिन्नत परीक्षण
test of homogeneity—समागता परीक्षण	union—संगम
testing of hypothesis—परिकल्पना की जाँच	unit—मानक, इकाई
theorem—प्रमेय	universe (population)—समष्टि
tosses—उत्क्षेपण	unknown—अज्ञात
treatment—उपचार	variance—प्रसरण
two dimensional random variable—द्वि-विमितीय मादृच्छिक चर	velocity—वेग
type—टंकन	vertical—ऊर्ध्व
	within group—समूहाम्यांतरिक